DTIC FILE COPY

AD-A211 608

Technical Report 839

Project A:
Improving the Selection, Classification, and Utilization of Army Enlisted Personnel

Assessing the Utility of MOS Performance Levels in Army Enlisted Occupations

Robert Sadacca
Human Resources Research Organization

Leonard A. White
U.S. Army Research Institute

John P. Campbell, Ani S. DiFazio, and Sheila R. Schultz
Human Resources Research Organization

May 1989

United States Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON ***Technical Director***	**JON W. BLADES** **COL, IN** **Commanding**

Research accomplished under contract for the Department of the Army

Human Resources Research Organization

Technical review by

Lawrence M. Hanser
Paul Twohig

Accesion For		
NTIS CRA&I		✓
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) HumRRO IR-PRD-88-15	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 839
6a. NAME OF PERFORMING ORGANIZATION Human Resources Research Organization 6b. OFFICE SYMBOL (If applicable) HumRRO	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) 1100 South Washington Street Alexandria, VA 22314-4499	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION -- 8b. OFFICE SYMBOL (If applicable) --	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-82-C-0531
8c. ADDRESS (City, State, and ZIP Code) --	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63731A; PROJECT NO. 792; TASK NO. 232; WORK UNIT ACCESSION NO. C1

11. TITLE (Include Security Classification)
Assessing the Utility of MOS Performance Levels in Army Enlisted Occupations

12. PERSONAL AUTHOR(S) Sadacca, Robert (HumRRO); White, Leonard A. (ARI); Campbell, John P.; DiFazio, Ani S.; Schultz, Sheila R. (HumRRO).

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 84/12 TO 87/09	1989, May	84

16. SUPPLEMENTARY NOTATION Prepared under Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel (Human Resources Research Organization, American (Continued)

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Entry-Level MOS; MOS performance; MOS utility; Personnel classification; Personnel selection; Project A; Scaling methods; Utility values

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

> Project A is the Army's long-term program to develop a complete personnel system for selecting and classifying all entry-level enlisted personnel. The utility measurement component deals with determining the relative utility to the Army of different levels of performance in entry-level Military Occupational Specialties (MOS). Because little research has been performed on such questions, exploratory work was done in a series of workshops with Army officers on how performance levels should be defined, what unit of measurement is appropriate for describing the relative value of differential job assignments across various MOS/performance level combinations, and how such values can best be estimated. Two scaling methods (pile placement and direct estimation) were selected and used to estimate utility values for 273 entry-level MOS. The research established that a coherent, reliable set of relative utility values can be derived at all performance levels for a wide variety of MOS.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Lawrence M. Hanser	22b. TELEPHONE (Include Area Code) (202) 274-8275 22c. OFFICE SYMBOL PERI-RS

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI Technical Report 839

16. SUPPLEMENTARY NOTATION (Continued)

Institutes for Research, Personnel Decision Research Institute, U.S. Army Research Institute)

Technical Report 839

Project A:
Improving the Selection, Classification, and
Utilization of Army Enlisted Personnel

Assessing the Utility of MOS Performance Levels in Army Enlisted Occupations

Robert Sadacca
Human Resources Research Organization

Leonard A. White
U.S. Army Research Institute

John P. Campbell, Ani S. DiFazio, and Sheila R. Schultz
Human Resources Research Organization

Selection and Classification Technical Area
Frances C. Grafton, Acting Chief

Manpower and Personnel Research Laboratory
Curtis Gilroy, Acting Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

May 1989

Army Project Number
2Q263731A792

Manpower and Personnel

FOREWORD

This document describes research conducted as part of the Army's large-scale manpower and personnel effort for improving the selection, classification, and utilization of enlisted personnel. The thrust for the project came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the current U.S. military selection/classification test battery) and other selection variables as predictors of training and performance.

The overall research effort is devoted to developing and validating Army Selection and Classification Measures and is referred to as "Project A." It is conducted under contract to the Selection and Classification Technical Area (SCTA) of the Manpower and Personnel Research Laboratory (MPRL) at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). This research supports the MPRL and SCTA mission to improve the Army's capability to select and classify its applicants for enlistment or reenlistment by ensuring that fair and valid measures are developed for evaluating applicant potential based on expected job performance and utility to the Army.

Project A was authorized through a letter, Deputy Chief of Staff for Operations, "Army Research Project to Validate the Predictive Value of the Armed Services Vocational Aptitude Battery," effective 19 November 1980; and a Memorandum, Assistant Secretary of Defense (MRA&L), "Enlistment Standards," effective 11 September 1980.

To ensure that Project A research achieves its full scientific potential and will be maximally useful to the Army, a governance advisory group composed of Army general officers, interservice scientists, and experts in personnel measurement, selection, and classification was established. Members of the last component provide guidance on technical aspects of the research, while general officer and interservice components oversee the entire effort; provide military judgment; periodically review the research progress, results, and plans; and coordinate within their commands. Members of General Officers' Advisory Group during the period covered by this report included MG W. G. O'Leksy (DMPM) (Chair), MG J. B. Allen, Jr. (DCSOPS), MG T. J. P. Jones (FORSCOM, DCSPER), MG G. Mallory (TRADOC, DCS-T), and BG P. M. Mallory (USAREUR, ADCSOPS). The General Officers' Advisory Group was briefed in May 1987 on the results of the concurrent validation, the preliminary results of the second-tour job analysis, and the plans for the longitudinal validation data collection. Members of Project A's Scientific Advisory Group (SAG) guide the technical quality of the research. During the period covered by this report, they included Drs. Philip Bobko, Thomas Cook, Milton Hakel (Chair), Lloyd Humphreys, Lawrence Johnson, Robert Linn, Mary Tenopyr, and Jay Uhlaner. The SAG was briefed in March 1987 on the status of the second-tour job analysis, the final resolution of utility measurement issues, and the reanalysis of the aptitude area composites. They were briefed in

September 1987 on the results of the utility and construct weighting research and the plans for second-tour criterion measurement.

A comprehensive ses of new selection/classification tests and job performance/training criteria have been developed and field tested, and the revised tests have been administered in a large-scale concurrent validation data collection effort and in longitudinal validation phases. The present report describes the work done to evaluate the relative utility of different levels of first-tour performance in Army enlisted occupations. Results from this and other Project A research activities will be used to link enlistment standards to required job performance standards and to more accurately assign soldiers to Army jobs.

Edgar M. Johnson

EDGAR M. JOHNSON
Technical Director

ASSESSING THE UTILITY OF MOS PERFORMANCE LEVELS IN ARMY ENLISTED OCCUPATIONS

EXECUTIVE SUMMARY

Requirement:

Project A is a long-term Army research program aimed at improving the system for selecting and classifying enlisted personnel. To develop an optimal classification system requires not only knowledge of how individual characteristics are related to job performance, but also information on the value to the Army of alternative decisions on job assignments. To provide this information, research was initiated to evaluate the relative utility of first-tour performance in Army enlisted occupations. This information can help decision makers maximize the organizational productivity that results from improved selection and classification procedures.

Procedure:

Sincen there has been little research on this topic, a series of exploratory workshops were conducted to develop a procedure that could be used to obtain performance utility values. In these workshops, Army officers considered in depth how to define different levels of performance in Army occupations, how to measure the value of performance at these different levels, and what the context for these utility judgments ought to be.

Two scaling methods developed in these initial workshops were then used to obtain performance utilities for 276 entry-level Army Military Occupational Specialties (MOS). Seventy-four field grade officers, representing a broad range of specialties, provided judgments of the relative value of different levels of first-tour performance in the 276 MOS. The performance levels were set at the 10th, 30th, 50th, 70th, and 90th percentiles, using the current recruit pool as the reference population.

Findings:

The officers' judgments for each job were averaged to provide mean MOS-specific utility values at the five performance levels. These mean values were quite stable across the different officer specialties and the two methods of scaling. Taken collectively, these results indicate that field grade officers share strikingly similar perceptions of the value to the Army of different levels of performance in enlisted occupations.

There is considerable variety in the pattern of judged utility values across performance levels for different occupations. In general, the benefit from exceptional performance was highest for mechanics and operators

of complex weapons systems. The relative cost of poor performance was greatest in the medical and maintenance specialties.

Utilization of Findings:

The most important finding in this first research effort is that the procedures developed here can be used to obtain reliable performance utilities for Army enlisted MOS. Differences in patterns of results across MOS reflect differences in the way that soldier performance during the first tour of duty contributes to organizational productivity.

A next direction for this research will be to examine the effect of using (or not using) this information on performance utility to guide decisions on personnel classification and job assignment. Assignment to maximize the utility of performance will incorporate not only MOS differences in the Army's capability to predict first-tour performance, but also the relationship of different levels of performance to overall Army effectiveness.

ASSESSING THE UTILITY OF MOS PERFORMANCE LEVELS IN ARMY ENLISTED OCCUPATIONS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ASSESSING THE UTILITY OF MOS PERFORMANCE LEVELS IN ARMY ENLISTED OCCUPATIONS

OVERVIEW OF PROJECT A

Project A is a comprehensive long-range research and development program the U.S. Army has undertaken to develop an improved system for selecting and classifying enlisted personnel. The Army's goal is to increase its effectiveness in matching first-tour enlisted manpower requirements with available personnel resources, through use of new and improved selection/classification tests that will validly predict carefully developed measures of job performance. The project addresses the Army's 675,000-person enlisted personnel system encompassing several hundred military occupations.

The program began in 1980, when the U.S. Army Research Institute (ARI) started planning the extensive research needed to develop the desired system. In 1982 ARI selected a consortium, led by Human Resources Research Organization (HumRRO) and including American Institutes for Research (AIR) and Personnel Decisions Research Institute (PDRI), to undertake the 9-year project. It is utilizing the services of 40 to 50 ARI and consortium researchers working collegially in a variety of professional specialties. The Project A objectives are to

- Validate <u>existing</u> selection measures against both existing and project-developed criteria (including both Army-wide job performance measures based on rating scales and direct hands-on measures of MOS-specific task performance).

- Develop and validate new selection and classification measures.

- Validate intermediate criteria, such as training performance, as predictors of later criteria, such as job performance, so that better informed decisions on reassignment and promotion can be made throughout a soldier's career.

- Determine the relative utility to the Army of different performance levels across MOS.

- Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making decisions.

The research design incorporates three main stages of data collection and analysis in an iterative progression of development, testing, evaluation, and further development of selection/classification instruments (predictors) and measures of job performance (criteria). In the first iteration, file data from fiscal years (FY) 1981/1982 were evaluated to explore relationships between scores of applicants on the Armed Services Vocational Aptitude Battery (ASVAB) and their later performance in training and their scores on first-tour Skill Qualification Tests (SQT).

For the ensuing research, 19 Military Occupational Specialties (MOS) were selected as a representative sample of the Army's 250+ entry-level MOS. The selection was based on an initial clustering of MOS derived from rated similarities of job content. These MOS account for about 45 percent of Army accessions and provide sample sizes large enough so that race and sex fairness can be empirically evaluated in most MOS.

In the second iteration, a Concurrent Validation design was executed with FY83/84 accessions. A "Preliminary Battery" of perceptual, spatial, temperament, interest, and biodata predictor measures was developed and tested with several thousand soldiers as they entered four MOS. The data from this sample were then used to refine the measures, with further exploration of content and format. The revised set of measures was field tested to assess reliabilities, "fakability," practice effects, and other factors. The resulting predictor battery, the "Trial Battery," was administered together with a comprehensive set of job performance indexes based on job knowledge tests, hands-on job samples, and performance rating measures, in the Concurrent Validation during the summer and fall of 1985. The results of the Concurrent Validation were used to form five performance constructs and to report to the Army incremental validities of the Trial Battery components over ASVAB predictors.

On the basis of testing experience, the "Trial Battery" was revised as the "Experimental Predictor Battery," which in turn is being administered in the third iteration, the Longitudinal Validation stage, which began in the late summer of 1986. All measures are being administered in a true predictive validity design. About 50,000 soldiers across 21 MOS are included in the FY86-87 administration and subsequent first-tour measurement. About 3,500 of these soldiers are expected to be available for second-tour performance measurement in FY91. Three MOS were added to the original 19 (19K, 29E, and 96B), and one of the original MOS was dropped (76W).

For administrative purposes, Project A is divided into five research tasks: Task 1, Validity Analyses, and Data Base Management; Task 2, Developing Predictors of Job Performance; Task 3, Developing Measures of School/Training Success; Task 4, Developing Measures of Army-Wide Performance; Task 5, Developing MOS-Specific Performance Measures.

Activities during the course of Project A research have been reported as follows: FY83, ARI Research Report 1347 and its Technical Appendix, ARI Research Note 83-37; FY84, ARI Research Report 1393 and two related reports, ARI Technical Report 660 and ARI Research Note 85-14; FY85, ARI Technical Report 746 and ARI Research Note 87-54; FY86, ARI Technical Report 792 and ARI Research Note 88-36; FY87, ARI Technical Report (in preparation) and ARI Research Note (in preparation); FY88, ARI Technical Report (in preparation). These reports list other publications on specific Project A activities. activities.

INTRODUCTION

This report describes the Project A research to determine the relative utility to the Army of different levels of performance in entry-level military occupational specialties (MOS). The main purpose of the utility measurement component in Project A is to provide information that will aid decision-makers in maximizing the payoff to the Army of improved selection and classification procedures.

Two major issues in developing and evaluating a personnel selection or classification system are how to maximize the gain to the organization from using the system and how to assess the net gain to the organization from using the new system versus not using it.

To answer such questions, at least three major elements are needed: (1) a model that portrays the relevant parameters in the decision-making process and specifies how they are interrelated, (2) a metric that can be used to represent the value of the outcomes that result from a particular course of action, and (3) a method for estimating the parameters of the model in the appropriate metric.

We know a fair amount about modeling personnel selection decisions (e.g., Cronbach & Gleser, 1965) and somewhat less, but still quite a bit, about modeling personnel classification decisions (e.g., Roulon, Tiedeman, Tatsuoka, & Langmuir, 1967). A great deal of effort by psychometricians and industrial psychologists has been put into developing and refining such models (cf. Cascio, 1982a). We are much less clear as to the metric in which the outcomes of a personnel selection or classification decision should be expressed.

The Utility Issue in Industrial Psychology

Although the steps described below have not occurred in a perfect chronological order, the progression of attempts by psychometricians and personnel researchers to portray the benefits of selection and classification acumen has been something like the following.

The validity coefficient, in the form of the product moment correlation between a predictor composite and a criterion composite, is the classic method by which the value of a selection program is represented. However, as is widely acknowledged, the correlation coefficient is a difficult metric to interpret. Early on, a number of transformations, such as the coefficient of determination (r^2_{xy}), the index of forecasting efficiency ($1-\sqrt{1-r^2_{xy}}$), and the standard error of prediction ($S_y\sqrt{1-r^2_{xy}}$), were suggested and found wanting. They still depended very heavily on the correlation coefficient itself, and cannot be interpreted directly in terms of benefits from decision making.

A more useful kind of transformation is represented by the various ways of using the bivariate distribution to construct decision tables. The Taylor-Russell tables (Taylor & Russell, 1939) are an example. With these transformations, the metric becomes the proportion of correct predictions that are made by one selection method versus another. One benefit of looking at selection payoff in terms of decision accuracy is that it illustrates quite clearly how even a small relationship between predictor and criterion can produce significant gains in the number of successful people selected if the selection ratio is very low and/or the variability in performance is high (e.g., base rate for success/failure = .50). However, to express the value of selection in these terms, the organization must define specific criterion categories (e.g., successful versus unsuccessful performance) and must view all the outcomes in a particular category as being equally valuable.

A new dimension was added by the classic work of Brogden (1946, 1949), who showed that if both the predictor and the criterion measures had interval properties and if the relationship between them was linear, then the correlation coefficient is linearly related to the gain in performance in the selected group. Further, the gain, in standard criterion units, that will result from selection can be estimated using existing prediction (i.e., decision) models if a cutting score is set on the predictor. Brogden also argued that a desirable metric for performance and performance gain would be to determine the dollar value of variability in performance.

It remained for Cronbach and Gleser (1965) to add the consideration of selection costs and to portray the utility of selection benefits in terms of the dollar value of performance increases minus the costs of selection. Cronbach and Gleser also elaborated the utility formulation to include more complex selection modes (e.g., multiple hurdles) and made an attempt to formulate classification decisions in utility theory terms.

The application of this kind of utility/decision theory to selection and classification problems was hampered by the difficulty of estimating the variability of performance in dollars, which is a major parameter in the model. Recently, Schmidt, Hunter, McKenzie, and Muldrow (1979) proposed a rather simple solution in which supervisors are used as judges to scale individual performance in dollar terms via a magnitude estimation technique. Judges are asked to estimate the dollar payoff to the organization of performance at the 50th percentile and the 85th percentile for people in the job in question. The difference between the two estimates is taken as the standard deviation of individual performance in dollar terms (SD_y). So far, not much attention has been paid to the basis on which supervisors make such judgments although the value for SD_y is frequently between 40 and 60 percent of the annual salary for the position.

Cascio (1982b) has proposed another technique for estimating SD_y in dollars that also uses expert judgment and is tied explicitly to salary. Job analysis is used to determine the major task components of a job, their relative importance is determined by expert judgment, and a magnitude estimation technique is used to rate every person's performance on each task factor. Average total salary is apportioned to each factor in accordance

with its importance weight. Average performance is set equal to 1.0 and the resulting scale is multiplied by the proportion of salary designated for that factor. Performance differences have thus been converted to a dollar metric and the standard deviations of the aggregate differences are put into the Cronbach and Gleser equation.

Utility Judgments in the Military Context

Two principal factors make it difficult to apply the previous work on utility metrics and utility estimation to the Army context. First, compensation practices are quite different in the Army in comparison with the civilian sector. Salaries do not differ by MOS and thus cannot be used as an index of the job's relative worth to the organization. Second, industrial firms are in business to provide products or services so as to maximize profit, whereas the Army's overall mission is to be prepared to defend the United States against military threats that everyone hopes will never come; it is difficult to try to put a monetary value on success or failure or to even think of the utility of jobs in terms of their monetary benefit.

While dollars may not be an appropriate metric with which to evaluate a new Army classification system aimed at maximizing preparedness for catastrophic events, resources are not unlimited. Choices among alternative personnel practices must be made, whether or not there is an explicit utility metric on which to make comparisons.

The Air Force Procedure

One operational answer to the evaluation problem is the system currently in use in the U.S. Air Force. Entry-level assignments in the Air Force are made by the PROMIS selection and classification system (Ward, Haney, Hendrix, & Pina, 1978). In brief terms, the individual assignment is a function of the following five parameters:

1) The level of predicted training success, using the ASVAB and other applicant information as predictors.

2) The individual's job preferences.

3) The rate at which the targeted quota for a job is currently being filled.

4) The rate at which the minority group targets for each job are being filled.

5) The scaled importance value of each combination of job holder Aptitude Level by Job Difficulty.

It is this last parameter that serves as the analog for a utility metric in the Air Force system. Previous scaling research using expert judges has produced an overall scale value for the relative importance of each combination of job difficulty (as determined by expert judgment) and

the aptitude level of a job holder (as determined by ASVAB scores). In general, the greater the job difficulty or the higher the aptitude level of the individual, the higher the value of that personnel assignment. However, the prediction surface that relates the aptitude level/difficulty level combination to assignment value is not a linear plane.

The approach of Project A to the problem is similar but not the same. Instead of scaling the relative importance of job difficulty X aptitude level combinations, the focus in Project A has been on assessing the differential value, or payoff, from MOS X predicted performance level combinations.

Specific Utility Issues for Project A

The overall objective of Project A is to produce the information necessary to develop a functional personnel classification system for all enlisted personnel. The objectives of its companion research endeavor, Project B, are to develop the necessary algorithms for relating labor supply forecasts, applicant information, and forecasts of system needs in an assignment system that uses Project A data in an optimal fashion. That is, whatever the increments in selection and classification validity produced by procedures developed in Project A, the Project B systems should allow investigation of how to maximize the benefit from using the new procedures.

Within this context, the utility problem for Project A becomes one of assigning utility values to MOS X performance level combinations. That is, if it is true that personnel assignments will differ in value to the Army depending on the specific MOS to which an assignment is made and on the level at which an individual will perform in that MOS, then the value of a classification strategy that has a validity significantly greater than zero will increase to the extent that the differential values (utilities) can be estimated and made a part of the assignment system.

For Project A the problem of estimating such utility values breaks down into a number of specific questions.

First, how should performance levels be defined? Should it be in terms of some general performance dimension that is left unspecified and is defined only in terms of relative level (e.g., percentiles)? Should a general performance dimension be explicitly defined, perhaps with behavioral anchors developed via critical incident methodology? Should individual performance components be defined and then explicitly weighted for combination into a total score? All of these are possibilities and a specific research question concerns how performance levels should be defined and described in the MOS X performance level combinations.

Second, what is the most appropriate metric for describing the relative value, or utility, of differential assignments across MOS/performance level combinations? Previous work in personnel psychology has been linked almost exclusively to a dollar metric and has tried to estimate the variability in payoff from people at different performance levels in dollar terms, but only in a selection context. Estimating differential payoff from a systemwide

classification system remains unexplored. Since the dollar metric appears to be inappropriate for the Army context and because there is little previous work on applying utility theory to personnel classification, the metric question for Project A is a very difficult one. It suggested an exploratory approach.

Third, assuming the question of the metric is resolved, the specific method(s) to be used for estimating differential assignment utility in the appropriate metric must then be considered. Only two options seem even possible. In the first, it might be possible to relate the performance of individuals or units to some kind of "bottom line" measure that Army management would consider an appropriate metric. For example, realistic field exercises could be used to determine the relationships of individual performance measures to the performance of a unit in a simulated engagement. The difficulties with this approach revolve around the expense of collecting such data, the necessity of having such exercises for each MOS, and the need to equate scores in some way across MOS.

A second alternative is to turn to scaling technology and use expert judges to estimate the relative value of differential personnel assignments. There are a variety of scaling models and scaling techniques from which to choose and a major problem would be to choose the procedure that is feasible, makes the best use of the information held by the judges, and provides sufficient internal validity information to generate confidence and acceptability for the scale values.

Since the above questions are difficult ones and have been largely unresearched in the past, the plan that was developed for addressing them was exploratory in nature. Its goal was to proceed from a very broad consideration of a number of methods to a focus on a procedure that is valid, feasible, and acceptable to the Army.

GENERAL APPROACH

Phase one consisted of a series of seven small group workshops with Army officers (Sadacca & Campbell, 1985). The workshops were designed to explore a number of issues pertaining to utility metrics, utility estimation, and the definition of performance levels. Each workshop was divided into a period for trying out prototypic judgment tasks and a period for open-ended discussion of issues.

Although the atmosphere was informal and the participants were free to bring up any questions or issues they wished, the following questions were used to guide the discussions:

1) How shall measures of performance be weighted and overall performance defined?

2) What kinds of scaling judgments can officers reasonably be asked to make?

3) Are there major scenario effects on performance factor weights and utility judgments?

4) In what metric should the utility of enlisted personnel assignments be expressed?

5) What is the form of the relationship between performance and utility within MOS?

6) Who will make the best judges for the final scaling?

The prototypic judgment tasks that were tried out in phase one were of the following general nature:

1) Assignment of importance weights to performance factors.

2) Rank ordering of overall utility of MOS by Performance Level combinations when performance was defined in percentile terms.

3) Ratio judgments of comparative utility for different MOS by Performance Level combinations.

The specific reactions of each participant to the sample scaling tasks were also used as items for general discussion.

The second phase of the research was devoted to solving the practical problems of assigning utilities to performance levels in the broad array of entry-level MOS. Additional workshops were conducted to try out various scaling methods and to prepare for the third phase, in which the selected scaling methods were applied to entry-level MOS and within MOS performance levels.

THE EXPLORATORY WORKSHOPS

In this first phase there was no vigorous testing of hypotheses, no experimental design or testing for statistical significance. If something didn't seem to work it was dropped or modified; if something else was suggested it was tried out. The overall intent was to determine what was possible, before being concerned with how to do it most effectively.

Workshop 1

A critical initial concern was whether Army officers would be willing to make evaluative judgments comparing the utility of enlisted soldiers in different MOS. Officers might, for example, argue that all military jobs are essential, and that is does not make sense to say that the soldier who transports the ammunition has any less utility than the soldier who fires the weapon, or the soldier who treats the wounded, or the soldier who prepares meals.

Another concern was what military situation or scenario should be used as the context in which utility was to be judged. It seemed very reasonable to believe that the utility to the Army of different military jobs, and performance levels within those jobs, would vary as a function of the stipulated military situation.

A third concern centered on what considerations enter into utility judgments made by Army officers. When evaluating a soldier's utility, what contributions to mission accomplishment do officers emphasize?

To get an initial understanding of the first two issues, it was decided not to provide any military context for making the utility judgments to officers attending the first workshop (six field grade officers from the Army Research Institute). The objective was to find out whether they would evoke their own military context for the judgments, and if so, what context they would choose. To assess the reasonableness of making any utility-type judgments, the only scaling task utilized was simply to ask for a rank ordering of MOS/performance level combinations, rather than for more sophisticated judgments that could yield an interval or ratio utility scale.

After a brief introduction to Project A and a discussion of the concept of job performance utility, the six officers were given the task of rank ordering a set of 57 enlisted MOS/performance level combinations from the set of 19 Project A MOS involved in the Concurrent Validation testing phase. To facilitate their judgments, they were provided with a separate listing of summary job descriptions for each MOS.

Perhaps the most important result was simply that the officers were willing to do the task. They did not argue that it was an unreasonable one and seemed to undertake the task quite seriously and carefully.

Another significant result emerged in the post task discussion: Independently, each of the six officers had chosen the same scenario -- that of a European conflict -- as the context in which they had rank ordered the utility of the MOS/performance level combinations. In the discussion period, the officers expressed the opinion that the Army's principal current mission is to ready itself for such a possibility. They also agreed that had we used a peacetime or a different wartime context, their utility rankings would most likely have been different. However, in their opinion, even if we used a peacetime scenario it should be one that emphasized training and other readiness activities geared toward the outbreak of hostilities in Europe.

The rank order intercorrelations among the officers were computed across the 57 MOS/performance level combinations. These correlations ranged from .29 to .90 with an average of .69. These results were heartening, since they indicated that quite reliable (.95 or above) average utility ranks could be obtained by using 10 or more judges. The results also indicated that there may be a fairly common frame of reference among Army officers in their evaluation of MOS/performance level utilities.

In the manner of a subjective expected utility model, the officers were next asked to evaluate the relative priority of eight outcomes (e.g., enhanced readiness, enhanced performance of supporting Army units) of a military engagement that could result from effective performance of enlisted personnel in that situation. The eight outcomes were chosen by the research staff without regard to any official Army doctrine. The military scenario (chosen before the workshop began) was one describing the outbreak of hostilities in Europe that had been used previously in Project A activities.

In the first variant, the officers first rank ordered the eight outcomes. Then, assigning ten points to the lowest ranked outcome, they assigned points to the remaining outcomes in accordance with the perceived ratio of their importance to the lowest ranked outcome. For a second method, the officers were presented the eight outcomes in a paired-comparison format; for each possible pair of outcomes, their task was to divide 100 points between the outcomes in a manner that reflected the outcomes' relative importance in the given military situation.

The officers distinctly did not like the paired-comparison format, feeling that it was like a test of their consistency in assigning importance points.

In the discussion period, the officers indicated that dollar cost considerations had no place on a battlefield, that losing or even winning a war could not be evaluated in dollar terms. They further indicated that the costs of training and equipping soldiers did not enter into their rankings of MOS/performance level utility.

In response to the question whether of judges should evaluate MOS/performance levels against separate utility dimensions, the officers expressed a clear preference for making one overall utility rating. They also felt that the description of the MOS/performance levels should be kept general rather than made more specific.

<u>Workshops 2 and 3</u>

The second and third workshops were scheduled back-to-back on successive days, with the intent of using the same stimulus materials and judgment tasks in both workshops. However, discussions with the officers in the second workshop led to changes in the procedures used the next day.

One such change involved the scenario used to describe the military context for the utility judgments. Discussions with the six field grade officers in the second workshop indicated that their utility ratings might well have been influenced by the type of unit to which they imagined themselves assigned. Furthermore, they might have been responding differentially to the "rugged, hilly and wooded" terrain description. One officer, for example, reported that he had downgraded the utility of armor crewmen because of the more limited use of tanks in that setting, while other officers reported that they had nevertheless assigned very high utility values to the armor crewman MOS.

The officers suggested keeping the scenario(s) free of specific details that would favor one MOS at the expense of another. The references in the scenario to the specific terrain and weather conditions were therefore deleted from the wartime scenario used in the third and subsequent workshops. Moreover, the military unit of concern was made the entire Corps, rather than an unspecified unit within the Corps. (See Figure 1.)

In both the second and third workshops, verbal descriptions of MOS/performance level descriptions were used. The descriptions were the same as those used in the first workshop, with one exception: The overall performance scale was changed from one which was behaviorally anchored to one expressed in percentiles (see Figure 2). This change was made in recognition of the difficulty of assigning performance-based anchors that would be comparable across MOS in the absence of actual performance data.

In the second and third workshops, in addition to rank ordering the described soldiers, the participating officers were asked to assess the relative utility of each of the soldiers in comparison to one particular or standard soldier whose utility was arbitrarily set at 100. The officers compared each of the 56 remaining soldiers in turn to the standard soldier and assigned a proportionate utility value to each, given that the standard soldier's value was set at 100. Two standard soldiers were used: the 90 percentile Infantryman (MOS 11B) and the 50th percentile Ammunition Specialist (MOS 55B). These two MOS/performance level combinations were, respectively, rank ordered very high and near the median by the first workshop officers. The officers were allowed to assign zero utility values or even negative values if they thought the soldier described would detract from mission accomplishment.

The average interjudge correlations and correlations between like utility measures across workshops were sufficiently high to suggest that very reliable average rank and/or ratio scale values could be obtained using about ten judges. The high intercorrelations among the different measures suggested that the final utility scale values (with appropriate transformations) might be fairly similar across measurement methods.

It was also apparent that, on the average, the combat MOS received higher utilities than the noncombat MOS at all three performance percentiles.

Differences in average scale values between the 90th and 50th percentile soldiers and between the 50th and 10th percentile soldiers suggested some important nonlinear relationships between performance and utility, which many investigators assume to be linear. Discussions with the officers in the workshops supported the nonlinear view.

In the discussion following the judgment tasks, the participants showed clear preference for the 90th percentile Infantryman rather than the 50th percentile Ammunition Specialist as an anchor, in part because they considered it easier to scale other MOS between the 0 and 100 points, and in part because Infantryman is the most common and best known Army MOS.

WARTIME SCENARIO: FIRST AND SECOND WORKSHOPS

Your unit is assigned to a U.S. Corps in Europe. Hostilities have broken out and the Corps combat units are engaged. The Corps 'mission is to defend, then re-establish, the host country's border. Pockets of enemy airborne/helicopter and guerilla elements are operating throughout the Corps sector area. The Corps maneuver terrain is rugged, hilly, and wooded, and weather is expected to be wet and cold. Limited initial and reactive chemical strikes have been employed but nuclear strikes have not been initiated. Air parity does exist.

WARTIME SCENARIO: THIRD - SEVENTH WORKSHOPS

Hostilities have broken out in Europe and your Corps' combat units are engaged. Your Corps' mission is to defend, then re-establish, the host country's border. Pockets of enemy airborne/heliborne and guerilla elements are operating throughout the Corps sector area. Limited initial and reactive chemicals strikes have been employed but nuclear strikes have not been initiated. Air parity does exist.

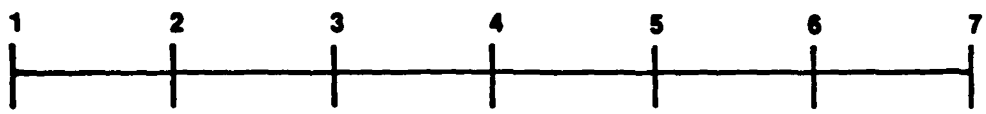
PEACETIME SCENARIO: FOURTH - SEVENTH WORKSHOPS

Europe is in the peacetime condition currently prevailing there. Your Corps' mission is to defend and maintain the host country's border should war break out. The potential enemy approximates a combined arms army and has nuclear and chemical capability. Air parity does exist. The Corps has personnel and equipment sufficient to make its mission capable for training and evaluation. The training cycle includes periodic field exercises, command and maintenance inspections, ARTEP evaluations, and individual soldier training/SQT testing.

Figure 1. Scenarios used in exploratory utility workshops.

INITIAL

Overall Soldier Effectiveness:



BELOW STANDARDS

Performs poorly in important effectiveness areas; does not meet standards and expectations for adequate soldier performance.

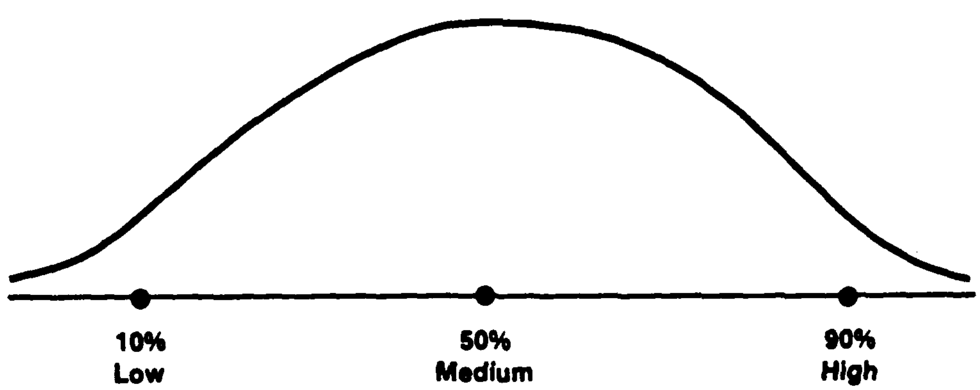
ADEQUATE/MID-RANGE

Adequately performs in important effectiveness areas; meets standards and expectations for adequate soldier performance.

SUPERIOR

Performs excellently in all or almost all effectiveness areas; exceeds standards and expectations for soldier performance.

CURRENT



OVERALL PERFORMANCE IN MOS

NOTE: 10th percentile indicates low overall performance and 90th percentile indicates high performance.

Figure 2. Description of soldier performance.

When asked what major factors they considered in assigning utilities to the MOS/performance combinations for the wartime scenario given, the officers indicated that potential contribution to unit survival and usefulness in replacing troop losses were foremost. This was consistent with the ratings given by the Workshop 1 officers of the relative importance of various outcomes.

When asked how general or specific the descriptions of the MOS/performance levels should be, the workshop participants said that most officers think in terms of top, bottom, and mid-level enlisted personnel. That is, a soldier is either good, poor, or somewhere in the middle. They felt that very general performance descriptions would best capture this outlook.

Workshops 4 and 5

The fourth and fifth workshops were conducted for the most part with the field grade officers who had participated in the first and third workshops.

The officers at both workshops were asked to follow new procedures that had not been tried out. Using the same wartime scenario and the 57 MOS/performance level combinations used in the third workshop, the officers were asked to judge 228 pairs of MOS/performance level combinations. The judgments were of the form:

> (______) soldiers of MOS/performance level combination 1 are equal in overall value to the Corps in the wartime military situation as (_______) soldiers of performance level combination 2.

The judgmental task was to fill in the two blanks with numbers that would make the two types of soldiers equal in value. For example, if the two MOS/performance level combinations were 90th percentile Utility Helicopter Repairer (MOS 67N) and 50th percentile Combat Engineer (MOS 12B), an officer might judge that seven of one type would be worth five of the other. The officers were allowed to put in any number they liked in order to make the two groups of soldiers equal in worth.

The 228 pairs of MOS/performance level combinations consisted of two types: (a) 57 pairs in which each pair member was from the same MOS but at a different performance level, that is, 10th, 50th, or 90th percentile (19 MOS x 3 pairs -- 10-50, 10-90, and 50-90); and (b) 171 pairs in which each pair member was from a different one of the 19 MOS, with one performance level for each MOS (19 x 18/2 = 171). The 228 pairs were randomized and then presented in the same order to all judges.

Scale values for each of the 19 MOS/performance level combinations making up the 171 judgmental pairs were calculated using a ratio scaling procedure described by Torgerson (1958, pp. 105-112). This procedure results in a set of scale values whose geometric mean is equal to 1.0.

Table 1 presents the average of the officers' scale values obtained for the 57 MOS/performance level combinations using the paired-comparison ratio scaling technique described above. Consistent with earlier findings, the combat MOS generally have higher utility ratings at each of the three performance levels (10th, 50th, and 90th percentile) than the noncombat MOS. However, the difference in utility scale values within an MOS from the 90th to 50th percentile performance level is greater for all 19 MOS than the difference in utility scale values from the 50th to 10th percentile performance level. This is especially evident for the combat MOS which, on the average, showed the greatest declines in utility values from the 90th to 50th percentile performance levels.

The inconsistency of these results with those cited earlier may be attributable more to the scaling method used than to the sample of officers involved, since the officers whose judgments were pooled to arrive at the Workshop 5 scale values overlapped considerably with the officers in Workshop 3.

The average interjudge correlation between the scale values of the eight officers taken across the 57 combinations was .61. This value, though not as high as that obtained for the scaling methods tried out in Workshop 3, was considered encouraging enough to try out the scaling method again in Workshops 6 and 7.

As five of the six officers in Workshop 5 had rank ordered the 57 MOS/performance level combinations using the same wartime scenario as in Workshop 3, one and one-half months earlier, it was of interest to determine how reliable their average rankings were. The correlation between the first and second average rankings by the five officers across the 57 combinations was .98.

Another indication of the stability of the average rankings is the average interjudge correlation obtained among the rank orders of the six officers. The obtained average, .79, is slightly [illegible] than the average obtained for Workshop 3 (.75). Both average interjudge correlations indicate that the average rank ordering based on 10 judges would probably have a reliability of .95 or better.

After the six officers in Workshop 5 finished scaling the MOS/performance level combinations, they were asked to re-rank the 57 combination cards under a peacetime scenario (see Figure 1). The peacetime scenario was set in Europe under current conditions and emphasized maintaining force readiness. Table 2 shows the MOS/performance level combinations having differences in average assigned rank of 10 or more under the wartime versus peacetime scenarios. The trend in the data from the six officers is clear: Low-performance-level combat troops are ranked higher in wartime than peacetime, while high-performance-level support personnel are ranked lower in wartime than peacetime.

Table 1

Scale Values of MOS/Performance Level Hypothetical Soldiers (50th Percentile Infantryman = 1.0; n = 8, Workshops 4 and 5)

MOS	Percentile			Scale Difference	
	10	50	90	(90-50)	(50-10)
Administrative Specialist (71)	.10	.23	.46	.23	.13
Ammunition Specialist (55B)	.17	.49	1.01	.52	.32
Carpentry/Masonry Specialist (51B)	.09	.21	.43	.22	.12
Chemical Operations Specialist (54E)	.26	.70	1.51	.81	.44
Food Service Specialist (94B)	.10	.23	.53	.20	.13
Light Wheel Vehicle Mechanic (63B)	.16	.43	.75	.32	.27
Medical Specialist (91B)	.21	.58	1.29	.71	.37
Military Police (95B)	.17	.34	.66	.32	.17
Motor Transport Operator (64C)	.12	.37	.68	.31	.25
Petrol. Supply Specialist (76W)	.13	.31	.71	.40	.18
Single Channel Radio Operator (05C)	.15	.41	.91	.50	.26
TOW/Dragon Repairer (27E)	.23	.64	1.26	.62	.41
Unit Supply Specialist (76Y)	.08	.23	.45	.22	.15
Util. Heli. Repairer (67N)	.17	.52	1.06	.54	.35
Noncombat MOS Average				.42	.25
Infantryman (11B)	.34	1.00	2.01	1.01	.66
Armor Crewman (19E/K)	.42	1.28	2.71	1.43	.86
Cannon Crewman (13B)	.29	.75	1.53	.78	.46
MANPADS Crewman (16S)	.27	.72	1.26	.54	.45
Combat Engineer (12B)	.26	.72	1.46	.74	.46
Combat MOS Average				.90	.58

Table 2

MOS/Performance Level Hypothetical Soldiers With Large Mean Wartime vs. Peacetime Differences in Rank Order (n = 6, Workshop 5)

		Mean Rank	
MOS/Performance Level	MOS	Wartime	Peacetime
Wartime Higher Than Peacetime			
Cannon Crewman, 10th percentile	13B	29	39
Cannon Crewman, 50th percentile		10	20
Chemical Opers Spec. 10th percentile	54E	35	48
Infantryman, 10th percentile	11B	25	40
Infantryman, 50th percentile		10	20
Armor Crewman, 10th percentile	19E/K	25	37
MANPADS Crewman, 10th percentile	16S	31	42
Peacetime Higher Than Wartime			
Administrative Spec, 10th percentile	71L	56	45
Administrative Spec, 50th percentile		46	28
Administrative Spec, 90th percentile		36	17
Carpentry/Masonry Spec, 50th percentile	51B	50	39
Carpentry/Masonry Spec, 90th percentile		41	26
Food Service Spec, 50th percentile	94B	41	25
Food Service Spec, 90th percentile		30	12
Unit Supply Spec, 90th percentile	76Y	28	14

The differences in average utility ranks found in Table 2 are certainly not surprising. They raise the question of how a computerized selection and assignment procedure can best use utilities if such utilities are in some part a function of the context in which the judgments of utility are made. It may be necessary to use utilities obtained through a number of scenarios or to decide upon one particular scenario as the context for the utility judgments. On the other hand, if the differences are not large, there may not be a significant difference in the recommended assignments to Army jobs using utilities obtained under different scenarios. The correlation across the 57 combinations of the average rank assigned by the six officers under the wartime and peacetime scenarios was .85. Computer simulations using different utility values and realistic operational constraints may eventually be needed to determine the practical significance of scenario differences.

After the officers had completed the judgmental tasks, a number of utility issues were discussed. The officers reported being concerned, when using the paired-comparison ratio scaling method, that they were being inconsistent in assigning numbers across the judgmental pairs of MOS/performance level combinations. They were assured that inconsistency could be expected within that type of judgment series. (The instructions were later modified in Workshops 6 and 7 to stress that it was not necessary to strive for consistency in making these kinds of judgments.)

When asked what MOS/performance level soldiers might best be used as a standard or unit in measuring the utility of other soldiers, the officers agreed that the 50th percentile Infantryman would be the best choice. They felt that not only are there more Infantrymen than soldiers in any other MOS, but that officers in general have a good understanding of what an average Infantryman is like and what he can do.

The officers were also asked what their reaction would be to expressing the differential worth or utility of soldiers in terms of dollars. They reacted very negatively to this concept, citing possible adverse political consequences as well as internal Army morale problems if dollar figures were placed on soldiers' worth.

<u>Workshops 6 and 7</u>

When the officers in Workshops 6 and 7, which were held in Europe, were asked the same question concerning the use of a utility dollar metric their reaction was, if anything, even more strongly negative. They, like the officers in earlier workshops, agreed that the 50th percentile Infantryman would make the best standard against which the utility of soldiers in other MOS/performance level combinations could be judged.

Thirteen officers attended Workshops 6 and 7. All were captains and majors, while the earlier workshop participants all had been majors and lieutenant colonels. The consistency of the opinions expressed by the officers in the discussion periods, despite the differences in grade levels and locations, suggests that Army officers have a fairly well-shared frame of reference.

This common viewpoint was also reflected in the results of the analyses of the workshop data. The Workshops 6 and 7 participants were asked to make essentially the same types of judgments made by earlier workshop participants. However, this time they judged the utility of 95 MOS/performance level combinations (5 performance levels--10th, 30th, 50th, 70th, and 90th percentile--for each of the 19 MOS) instead of 57 combinations. The correlation was .94 across the average paired-comparison ratio scale values of the 57 combinations that were common between Workshops 4 and 5 (majors and lieutenant colonels) and Workshops 6 and 7 (captains and majors).

The means of the rank orders assigned the 95 MOS/performance level combinations under the war and peacetime scenarios by the 13 officers are shown in Table 3. The MOS in the table have been placed in three groups based on comparative rankings. The first group contains mostly combat MOS.

Table 3

Mean Rank Order of MOS/Performance Level Combinations Under Wartime and Peacetime Scenarios (n = 13, Workshops 6 and 7)

		Performance Percentile				
MOS		10	30	50	70	90
Ranked Higher in Wartime Scenario						
Infantryman	W	64	43	23	13	6
(11B)	P	83	69	47	33	17
Armor Crewman	W	66	48	25	15	7
(19E/K)	P	83	67	47	32	16
Cannon Crewman	W	68	48	27	17	9
(13B)	P	83	68	51	32	16
Chemical Operations	W	73	57	40	24	14
Specialist (54E)	P	86	69	52	39	19
Single Channel Radio Operator	W	77	60	41	26	14
(05C)	P	84	72	53	36	15
Combat Engineer	W	72	52	32	27	14
(12B)	P	87	68	49	34	19
MANPADS Crewman	W	74	53	37	24	15
(16S)	P	85	68	54	35	18
Ranked Higher in Peacetime Scenario						
Administrative Specialist	W	88	77	68	57	41
(71L)	P	80	57	40	25	7
Unit Supply Specialist	W	82	73	54	42	27
(76Y)	P	76	61	39	22	7
Light Wheel Veh Mech	W	79	63	48	33	23
(63B)	P	79	61	42	27	10
Food Service Specialist	W	83	70	56	46	35
(94B)	P	81	60	44	28	12
Carpentry/Masonry Specialist	W	91	84	75	67	56
(51B)	P	84	65	50	34	20
Mixed						
Medical Specialist	W	74	57	45	26	15
(91B)	P	82	61	42	23	7
TOW/Dragon Repairer	W	80	62	50	35	27
(27E)	P	83	67	48	33	15
Utility Helicopter Repairer	W	76	61	45	34	23
(67N)	P	81	65	43	28	18
Motor Transport Operator	W	80	63	52	39	33
(64C)	P	82	65	45	29	13
Military Police	W	81	66	53	41	28
(95B)	P	83	67	45	31	11
Petrol Supply Specialist	W	79	64	50	32	20
(76W)	P	82	63	48	30	12
Ammo Specialist	W	78	65	48	33	22
(55B)	P	84	72	55	37	19

All the MOS/performance level combinations involving these MOS had higher average rank orders under the wartime than under the peacetime scenario. In the second group of MOS all the MOS/performance level combinations were ranked higher under the peacetime than the wartime scenario. In the third group of MOS the average rank orders of the MOS/performance level combinations were all higher under peacetime than wartime at the upper levels of performance, but were all lower under peacetime than wartime at the lower levels of performance. Soldiers in these MOS generally have a higher probability of being in a combat situation than soldiers in the second group of MOS.

These data were consistent with the Workshop 4 and 5 findings and the statements made during the discussion periods: Soldiers at low performance levels who are likely to be involved in combat are assigned relatively higher utility under a wartime scenario, while soldiers at high performance levels who are unlikely to be involved in combat are assigned relatively higher utility under a peacetime scenario. However, since the correlation across the 95 combinations of the utility values under the two scenarios may be quite high (the correlation of average rank orders was .83 in the Workshops 6 and 7 data and .85 for the comparable Workshops 4 and 5 data), the simulations may well result in relatively minor scenario-derived differences.

In Workshops 6 and 7, 12 of the officers scaled the 95 combinations in two ways. One method was the paired-comparison ratio procedure used by the Workshops 4 and 5 participants. They also scaled the 95 combinations using the subjective estimation procedure employed by the Workshops 3 and 4 participants. In this method one combination is given a utility value of 100 and the other combinations are assigned scale values that reflect their respective proportionate utilities; the combination assigned the value of 100 was the 90th percentile Infantryman. The scales obtained by the two methods were then transformed to scales in which the 50th percentile Infantryman had a utility value of 1.0.

Table 4 shows the scale values of the 95 MOS/performance level combinations obtained through using both methods. The utility scale values obtained from the two methods are quite similar at the lower performance levels. However, with the exception of the Infantryman and Armor Crewman MOS, the scale values for the higher performance levels obtained from the subjective estimation procedure are higher than those obtained using the paired-comparison ratio scaling technique.

Examination of the utilities assigned to the performance levels within MOS revealed that on the average, for both the combat and noncombat MOS, the subjective estimation utility values had a somewhat greater decline in the lower half of the performance levels (between the 50th and 10th percentiles) than in the upper half (between the 90th and 50th percentiles). The paired-comparison utility values, on the other hand, on the average had a somewhat greater decline in the upper half of the performance levels than in the lower half for both kinds of MOS.

Table 4

Mean Values of MOS/Performance Level Combinations Using Subjective Estimate and Paired-Comparison Ratio Scaling Techniques (n = 12, Workshops 6 and 7)

MOS	Method[a]	Performance Percentile 10	30	50	70	90
Administrative Specialist (71L)	SE	-.07	.29	.47	.74	.86
	PC	.09	.16	.24	.31	.45
Ammunition Specialist (55B)	SE	.12	.46	.69	.90	1.13
	PC	.12	.26	.38	.52	.73
Cannon Crewman (13B)	SE	.30	.69	.93	1.24	1.49
	PC	.24	.41	.64	.90	1.28
Carpentry/Masonry Specialist (51B)	SE	.00	.09	.37	.61	.80
	PC	.07	.11	.18	.24	.38
Chemical Operations Specialist (54E)	SE	.20	.53	.86	1.16	1.38
	PC	.16	.35	.48	.67	.96
Combat Engineer (12B)	SE	.20	.65	.96	1.22	1.52
	PC	.19	.38	.57	.77	1.05
Food Service Specialist (94B)	SE	.09	.33	.59	.83	1.04
	PC	.11	.18	.27	.38	.50
Infantryman (11B)	SE	.29	.71	1.00	1.30	1.58
	PC	.39	.63	1.00	1.53	2.18
Light Wheel Vehicle Mechanic (63B)	SE	.17	.51	.68	1.02	1.24
	PC	.13	.24	.37	.50	.65
Armor Crewman (19E/K)	SE	.40	.68	1.03	1.26	1.60
	PC	.25	.48	.73	1.14	1.63
MANPADS Crewman (16S)	SE	.19	.57	.83	1.09	1.38
	PC	.16	.31	.45	.65	.96
Medical Specialist (91B)	SE	.17	.48	.79	1.07	1.37
	PC	.15	.30	.42	.62	.95
Military Police (95B)	SE	.15	.47	.71	.97	1.20
	PC	.16	.26	.38	.52	.74
Motor Trans. Operator (64C)	SE	.06	.39	.59	.83	.97
	PC	.13	.21	.33	.43	.65
Petrol. Supply Specialist (76W)	SE	.16	.51	.72	.82	1.11
	PC	.13	.25	.39	.52	.78
Single Channel Radio Operator (05C)	SE	.13	.54	.77	1.09	1.30
	PC	.16	.26	.42	.53	.80
TOW/Dragon Repairer (27E)	SE	.10	.53	.74	.99	1.33
	PC	.16	.28	.43	.56	.78
Unit Supply Specialist (76Y)	SE	.08	.40	.60	.91	1.07
	PC	.12	.22	.34	.50	.69
Utility Helicopter Repairer (67W)	SE	.15	.49	.82	1.06	1.32
	PC	.17	.30	.43	.62	.90

[a] SE: Slightly greater decline in lower half than in upper for both combat and noncombat.

PC: Slightly greater decline in upper half than lower half for both combat and noncombat but somewhat larger for combat.

As in the case of the scenario differences, these scaling method differences may or may not have practical significance. The correlation between the mean values assigned the 95 combinations by the two methods was .91.

It is also of interest to note that in general the highest disagreement in assigning scale values occurred with high-performance-level noncombat MOS combinations, whereas the highest agreement in assigning scale values occurred with low-performance-level noncombat MOS combinations.

In general, however, as noted earlier, the Army officers seem to have a fairly common frame of reference. The median intercorrelations among the officers for the wartime rank orders and scaling values ranged from .76 to .80. Average scale values based upon the judgments of 10 or more officers should therefore have reliabilities of .95 or higher.

Summary Comment

Perhaps the most significant finding is that Army officers would be willing and able to assign differential utility values across MOS and performance levels. Perhaps the next most significant finding is that fairly stable scale values could be obtained from averaging across a relatively small number of officer/judges.

In addition, the scenario(s) used should be free of the detail that suggests greater or less utility for certain specific MOS. Utilities of soldiers in wartime should not be expressed in terms of dollars; an acceptable metric would be the utility of a 50th percentile Infantryman (his value for the survival of the unit and in replacing troop losses is much more readily apparent). Directions to the judges should be reassuring concerning inconsistencies that can possibly occur in a long series of judgments.

As discussed earlier, some of the problems identified (e.g., scenario effects) may have little practical significance in terms of how a computerized enlisted personnel selection and classification system would process Army applicants under operational constraints. Further research should examine, through sensitivity analyses and computer simulations, how differences in the utilities of MOS/performance level combinations affect system output.

TRYING OUT METHODS FOR ASSIGNING UTILITIES TO PERFORMANCE LEVELS

The second phase of utility scale development was devoted to developing the final procedures to be used in actually assigning utilities to performance levels in all entry-level MOS. Several inferences were made from the exploratory findings in the earlier workshops.

First, the apparent nonlinear relationships between utility and performance found in some MOS would necessitate obtaining judgments of the utility of at least five performance levels within each MOS. Five data

points would allow the derivation of a best fitting utility/performance curve with two inflection points (if necessary) within an MOS.

Second, the task of assigning utility scale values to at least five performance levels in 275 MOS was much too onerous to assign to any one judge. Some system for obtaining the judgments would need to be employed that allowed the task to be divided up among groups of judges, but that still allowed utilities to be reliably scaled both between and within MOS.

Third, the system used to obtain judgments from a group of judges could employ more than one scaling method. The high correlations between utility values obtained earlier from different scaling methods suggested that a combination of methods might allow the overall scaling task to be accomplished more efficiently than through using one method cnly. The goal was to place on the same ratio scale the utility values of at least 275 x 5, or 1,375, performance level/MOS combinations. (A ratio scale would permit utilities to be summed across individual MOS assignments in comparing selection/classification systems.)

Procedure

An additional 12 workshops were then conducted to try out various scaling methods and to prepare for the third phase of the research in which the selected scaling methods would be applied to all entry-level MOS and within MOS performance levels. These workshops, like the previous ones, were attended by small groups of Army field grade officers.

The methods tried out at the workshops were rank ordering, paired comparisons, a conjoint scaling procedure, the sorting or placement of MOS/performance level combinations into piles, and the direct estimation of ratio scale values using a standard MOS/performance level set at 100. Of these techniques, the latter two were the scaling procedures eventually selected.

Alternative Methods

The rank ordering task involved rank ordering a list of 135 MOS with all performance levels set at the 50th percentile. This method produced negative reactions from the workshop participants. They objected to the time it took to perform the rankings and to their inability to assign tied ranks under the method used. They felt that they did not know enough about all the MOS to make the fine discriminations called for in rank ordering. They also objected to the very task of rank ordering MOS, saying that all Army MOS were important. Though the Project A staff anticipated this latter reaction from some officers in earlier workshops, it was not exhibited until, as in this instance, participants were not asked for utility judgments of performance levels both within and between MOS.

Four issues were incorporated in the subsequent workshops in phase two. First, it was decided to use only scaling methods that allowed judges to report that two or more MOS performance levels combinations were approximately equal in utility. Second, the judges were offered the alternative

of not evaluating the utility of some MOS if they felt they did not know enough about the MOS to make informed judgments. Third, different performance levels were included within MOS, as well as between MOS, in the set of combinations to be judged.

The fourth change involved placing the judgments in a selection and classification context. That is, the officers were asked to judge the utility of predicted performance of Army applicants or recruits rather than actual performance of Army job incumbents (as had been done in the earlier workshops). Percentile levels were still used as in earlier workshops, but the percentiles were for predicted performance for the given MOS of all Army applicants or recruits. The judges were asked to assume that the performance percentiles given were accurate estimates of future on-the-job performance percentiles if the applicants or recruits were actually assigned to the MOS. After this adjustment was made, none of the judges in subsequent workshops objected to the basic concept of assigning differential utilities to various MOS/performance levels.

Two variants of the method of paired comparisons were also tried out using a limited number of MOS performance/level combinations. One involved judgments of number equality, as in Workshops 4 and 5. The other involved assigning 100 enlisted applicants with given predicted performance percentiles to pairs of MOS; for example, if there were 100 applicants who were at the 10th percentile for the job of illustrator (MOS 81E), and at the 30th percentile for the job of physical activities specialist (MOS 03C), how many of the 100 applicants should be assigned to each job?

Though both of these paired-comparison tasks called for complex judgments, the officers performed then readily. However, the methodology was time consuming, and would be even more so with larger numbers of MOS/performance level combinations to judge. Moreover, the officers felt they should be allowed to indicate that some applicants should not be selected at all. The judgment was subsequently shifted from predicted performance levels of applicants to that of recruits (selected applicants), thereby eliminating the "do not select" alternative. However, the judges were allowed to indicate that they thought a given recruit would have zero or negative utility for the Army if placed in an MOS where his or her predicted performance was low.

Related issues that arose were the field strength of Army units staffed with various MOS complements, and the possibility of potential troop losses if open warfare broke out. Some officers reported that they considered these factors in their utility evaluations of the applicants or recruits. To divorce both troop strength and troop replacements from utility/assignment decisions, judges in subsequent workshops were told that the field strength of all MOS was 70 percent and that the problem of compensating for troop losses was being handled by another part of the assignment algorithm and should not enter into their MOS utility judgments.

A conjoint scaling method was also tried out to determine whether it was possible to obtain MOS/performance level utility evaluations at the same time that weights were established from different components of performance.

Each of 16 MOS was paired with each other MOS in the set, at the same time that predicted percentile levels for 15 different pairs of performance factors were given. Although a conjoint procedure later proved effective for use in arriving at weights for combining performance factors into overall measures of MOS performance (see Sadacca, Campbell, White, & DiFazio, in preparation), the method tried here was much too difficult and time consuming for use in scaling large numbers of MOS/performance level combinations.

One method that did prove effective for making large numbers of scaling utility decisions was the pile placement method, in which judges sorted cards containing MOS/performance level combinations into piles, based upon their perceived utility or selection priority. Seven piles of predicted performance utility were used, ranging from negative through zero utility to high utility. The judges initially sorted 135 MOS/performance level combinations, then 210 combinations, and eventually 280 combinations, without complaining about the judgment burden.

Likewise, judges in the ratio judgment method, in which they evaluated MOS/performance level utilities in relationship to that of an 90th percentile Infantryman, judged 59 combinations without the task becoming burdensome.

Using data from one of the last workshops in phase two, separate analyses of variance were performed on the mean pile placements and ratio judgments given 59 combinations judged by the 12 officers using both methods. Remarkably similar *F* ratio patterns were obtained (see Table 5). For both scaling methods, there were large mean differences in assigned ratings for different predicted performance levels. Likewise, there were significant mean differences for rater and MOS, while the MOS x percentile level interaction was not significant in either analysis. The intraclass correlation reliability estimate for the pile placement procedure was .58 and the comparable coefficient for the direct ratio judgment was .65.

Table 5

Analysis of Variance of Sorting and Ratio Utility Values, Based on 59 Common Combinations (n = 12 Officers, Workshop 11)

		Sorting			Ratio		
Source	df	F	P> F	R^2	F	P> F	R^2
Model	69	14.76	.0001	.61	23.54	.0001	.72
Error	638						
TOTAL	707						
MOS	11	1.91	.0358		10.98	.0001	
Level	4	220.37	.0001		319.42	.0001	
MOS X Level	43	.61	.9771		.90	.6589	
Rater	11	8.17	.0001		17.01	.0001	

These results indicated that satisfactory reliabilities for mean utilities could be obtained by both methods, if the means were based upon 10 or more judges. The correlation between the mean utilities assigned by the 12 officers to the 59 common combinations using the two methods was .89. Though this intermethod correlation was not as high as might be desired, it shoud be increased by using MOS/performance level combinations with a greater range of utilities and by raising the reliability of both sets of scale values by an increase in the number of judges.

Methods Selected

In the light of all the information available from the first and second phase workshops, it was decided to use the pile placement and direct ratio estimation methods in the final determinations of the utilities of approximately 275 MOS x 5 performance level, or 1375 combinations. The pile placement method provided a means of reliably scaling the utility of large numbers of combinations in a reasonable time period, while the direct estimation method could be used to place a limited number of combinations on a reliable ratio scale having a meaningful zero point and the potential for assigning negative utilities to low predicted performance levels.

The procedures used to place the 1375 combinations on one utility scale are described in the next section, which also presents the results of the third and final phase of the utility scaling effort.

OBTAINING A COMPLETE SET OF UTILITY ESTIMATES

The goal of the exploratory workshops was to develop a scaling method(s) for obtaining utility functions for a large set of MOS, that to the maximum extent possible reflected the relative payoff to the Army of different levels of job performance. The results of these exploratory workshops were largely successful. First, for the Army jobs being considered in these workshops, utility scale values varied across MOS in a manner generally consistent with expectations. Second, the utility values assigned by the officers were sufficiently alike to indicate that fairly stable scale values could be obtained by averaging across officer judgments. Collectively, these results pointed to the feasibility of obtaining information on the relative value of performance in Army MOS that could be used to guide decisions in a personnel selection and classification system.

The next goal was to assign a utility to any predicted level of performance on any entry-level MOS. The obtained utility values could be used to (a) assess the net gain to the Army of using new selection/classification procedures, and (b) help guide classification algorithms in optimizing assignments of individual recruits.

Method

Design Issues. Observations made in earlier workshops of the amount of time it took field grade officers to place MOS/performance level combinations in piles on the basis of judged utility indicated that they could

judge 250 combinations in about 1 1/2 hours. Similar observations of the amount of time it took to directly judge the utility of combinations relative to the standard of a 90th percentile Infantryman indicated that the officers could judge 50 combinations in about 40 minutes. It was apparent that only a subset of the total number of MOS/performance level combinations could be presented to any one officer.

To place all utilities on a ratio scale, the project staff chose to use both the pile placement method and a direct judgment method. The pile placement method would be used to place the utilities on an interval scale and the direct estimation method would be used to develop a ratio scale for a target set of MOS x performance combinations against which the interval scale values could be calibrated.

To merge the utilities obtained from the two methods, the same officers should judge a common set of MOS/performance level combinations using both methods. Therefore, to adjust for differences among the samples of officers assigned particular subsets of MOS/performance level combinations in the pile sorting method, <u>all</u> judges were asked to judge one common set of 60 combinations using both methods.

Another issue concerned the number of performance levels within each MOS. Because of the large number of entry-level MOS, the number of performance levels within each MOS was restricted to five. This number still allowed for the derivation of a nonlinear function with inflection points when expressing utility as a function of the level of performance within each MOS.

<u>Selection and Grouping of Entry-Level MOS</u>. AR 611-201, <u>Enlisted Career Management Fields and Military Occupational Specialties</u> (October 1985), was used in the selection of 276 entry-level MOS. All MOS that listed Skill Level 1 duties and that required an ASVAB aptitude area score for assignment to the MOS were selected.

Consequently, there were 276 MOS times five levels or 1,380 MOS/performance level combinations to be judged separately. To make the scaling task more acceptable to the judges, seven sets of combinations were formed. The first set consisting of 12 MOS times five performance levels, or 60 combinations, was to be judged by all judges and would provide the basis for a common utility scale.

The Infantryman (11B) MOS was selected first because the utility of the 90th percentile Infantryman was to be used as the standard (set at 100) against which the utility of all other MOS/performance level combinations was to be compared in the direct judgment method. Judgment data from earlier workshops were used to identify 11 additional MOS that met the following three criteria: (a) No officers had refused to scale the MOS because of unfamiliarity with the MOS; (b) utility values for the 55 MOS/performance level combinations were evenly spread across the range of utilities assigned all MOS used in the workshops; and (c) extremely low or negative utility values were likely to be obtained for performance in some

of the MOS at the 10th percentile and high utility values were expected for some jobs at the 90th percentile.

The remaining 264 MOS were grouped into six subsets of 44 MOS each. The subsets were made comparable through a systematic, stratified assignment procedure. The MOS were first grouped in accordance with the results of a cluster analysis based upon judgments of job and task similarity (Hoffman, 1987). This analysis identified 23 MOS clusters. The MOS in each cluster were placed in numerical-alphabetical sequence. Then every sixth MOS was assigned to one of the six subsets, which were labeled Decks A, B, C, D, E, and F.

Each subset or deck contained 280 MOS/performance level combinations -- 12 common MOS plus 44 noncommon MOS and five performance levels for each job.[1] The combinations in each deck were randomized before being administered to the judges.

Sample of Officers Used as Judges. Data from the exploratory workshops indicated that about 10 officers would be needed to obtain an interjudge reliability of about .95 in utility judgments. To ensure that a total sample of 60 officers (10 officers x 6 decks) was obtained, utility workshops were held at six CONUS Army posts and in USAREUR. Altogether, 74 field grade officers attended the workshops. The 74 participants consisted of 14 Infantry, 21 Armor, 14 Other Combat (e.g., Artillery), 12 Combat Support, and 13 Combat Service Support officers. Most of the officers were majors; there were 54 majors, 13 lieutenant colonels, and 7 colonels among the participants.

The Utility Judgment Workshops

After a brief overview of Project A and description of the workshop agenda, the participants completed a Background Information Sheet, including items pertaining to grade, military specialty, current and previous positions, and years of service.

The workshop leader then gave a more detailed overview of the workshop and its purpose, and discussed assumptions that the participants were to use in making their judgments. The three critical assumptions are given below (the complete set of assumptions and copies of all the workshop briefing materials, instructions, and forms are supplied in Appendix A).

---

[1]Three noncommon MOS and their performance levels were subsequently dropped from the judgment sets since the three MOS were rescinded from operational status. Two of these MOS were in Deck C, the other in Deck E. The number of combinations in the six decks therefore ranged from 270 to 280. For convenience, the number 280 is used in the text and tables to indicate the number of MOS/performance level combinations in the decks.

1) The military context for which the utility of the recruits is being considered is as follows: The world is in a period of heightened tensions. There is an increasing probability that hostilities will break out in Europe, Asia, the Caribbean, Latin America, and Africa. The Army's mission is to support U.S. treaty obligations and to help defend the borders of allied and friendly nations. Some of the potential enemies have nuclear and chemical capability. Air parity does exist between allied forces and potential hostile nations. U.S. Army training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

2) The overall MOS performance measure for each MOS represents an optimally weighted (for that MOS) combination of several performance factors. Thus, recruits at the highest predicted performance level (90th percentile) in each MOS are more likely to be dependable, be proficient in MOS tasks, know the facts and procedures required to do their jobs, perform more effectively under adverse or difficult conditions, avoid disciplinary problems, provide support to fellow soldiers, and be more physically fit.

3) The predicted performance levels for the recruits are accurate. That is, the recruits will actually perform at the predicted levels.

Note that the judgment called for in both utility scaling methods was the relative value of recruits with different predicted performance levels for the entry level MOS. The decision to use predicted performance levels rather than actual performance levels was based upon the fact that, in application, a computerized enlisted personnel selection and classification system would be operating with predicted or estimated performance. (A major purpose of Project A is to improve the accuracy of that prediction.)

The participants then read a description of the <u>pile placement method</u>. Emphasis was placed on the definitions given the seven piles in which the judges were to place the MOS/performance level combinations:

- High positive utility would probably result if these recruits were placed in these MOS.

- Between moderate and high utility would probably result if these recruits were placed in these MOS.

- Moderate utility would probably result if these recruits were placed in these MOS.

- Between low and moderate utility would probably result if these recruits were placed in these MOS.

- Low positive utility would probably result if these recruits were placed in these MOS.

- Advantages of placing these recruits in these MOS would probably be equal to the disadvantages (expected utility = 0).

- Negative utility would probably result if these recruits were placed in these MOS. (Any positive contribution would probably be outweighed by problems associated with low levels of overall performance.)

Each MOS/performance level combination was printed on a separate card (see samples in Figure 3). On each card there was a short description of the MOS (Skill Level 1) taken from AR 611-201. The performance level, either the 10th, 30th, 50th, 70th, or 90th percentile, was also listed on the card. The instructions indicated that the percentiles were the predicted performance percentiles of recruits, if all recruits were rank ordered in terms of their predicted performance in the given MOS without regard to current cut-off scores. The instructions also allowed the judges to place in an eighth unrated pile any MOS/performance level combinations that they were not familiar enough with to evaluate. No restrictions were placed on the number of cards that could be placed in any one pile. The cards in each deck were thoroughly shuffled prior to the workshops. Decks were assigned to participants randomly.

Upon completing the pile placement method and a short break, the judges read the instructions for the direct judgment method (see Appendix A). After first reviewing the assumptions and re-familiarizing themselves with the 12 common MOS, the participants wrote the value, 100, on the 90th percentile Infantryman card, which was on top of the deck of 60 cards. The task of the judges was to assign a utility value to each of the remaining 59 MOS/performance level combinations, taking into consideration that a 90th percentile Infantryman had a utility of 100. Zero and negative utility values were permitted.

The judges wrote the assigned utilities directly on the cards. After they had gone through the deck once, they were instructed to arrange the cards in ascending numerical order and then go through the cards again and change any utility values that they felt were out of line with the others in terms of the ratios of the assigned utilities.

Reliability and Validity Analyses

Identification and Deletion of Atypical Judges

An initial question posed was whether some of the judges' responses were sufficiently atypical on *a priori* grounds to warrant excluding these judges from later analyses. For example, if any of the participants did not fully comprehend the task or its underlying assumptions, or if they were inattentive in accomplishing the task, then inclusion of their data could decrease the reliability and validity of the final scale values.

RADIO TELETYPE OPERATOR
(Radio TT Operator) (MOS 05C)

SUMMARY: Supervises or operates and installs radio teletypewriter and tape relay equipment in radio teletypewriter and tape relay tactical or administrative communications nets.

DUTIES: Operates radio teletype equipment to transmit and receive messages.

OVERALL EFFECTIVENESS: 70 percentile

HEAVY ANTIARMOR WEAPONS INFANTRYMAN
(MOS 11H)

SUMMARY: Leads or serves as member of heavy antiarmor crew-served weapons squad, section, or platoon employing heavy antiarmor crew-served weapons in offensive and defensive combat operations.

DUTIES: Assaults and destroys enemy tanks and armor vehicles, emplacements, weapons, and personnel with heavy antiarmor weapons (TOW).

OVERALL EFFECTIVENESS: 30 percentile

Figure 3. Samples of MOS/Performance Level combination cards

Four indexes were used to determine the degree of atypicality:

1) The number of times a judge assigned greater utilities to lower performance level recruits than to higher performance level recruits in the same MOS.

2) The median correlation of the utilities assigned by a judge across the MOS/performance level combinations with the utilities assigned by the other judges.

3) The mean utility assigned the MOS/performance level combinations by a judge. Unusually high or low mean values would indicate that the judge was assigning many of the combinations greater or lesser utilities than the other judges were.

4) The standard deviation of the utilities assigned the MOS/performance level combinations by a judge. Large or small standard deviations would indicate that the judge was assigning an unusually wide or narrow range of utilities to the combinations.

These indexes were calculated for both the pile placement and the direct judgment data from each judge. Frequency distributions of the eight sets of indexes were examined and the judges who had relatively extreme values were identified.

Of the eight indexes, those considered to be most indicative of atypicality were high numbers of inversions and/or low median correlations with other judges for either the pile placement or the direct judgment data. A "rule of thumb" was adopted that a judge had to have atypical values on at least one of these four indexes to be considered for exclusion.

On the basis of this rule, seven judges were excluded. Tables 6 and 7 show the frequency distributions for the number of within-MOS inversions.

Table 8 shows the frequency distributions of the median correlations between the judges. The median correlations for the pile placement data were obtained by forming a separate inter-correlation matrix for the judges assigned each deck. Here the correlation between any pair of judges was computed across their joint pile placements of the 280 combinations in their respective decks. The median correlations for the direct judgment data were obtained by forming an intercorrelation matrix for all judges. The correlation between any two judges was computed across the utility values they assigned the 60 common combinations. While over 90 percentile of the judges had median correlations above .50 for both types of judgments, six judges had median correlations of .50 or below for either the pile placement or the direct judgment method. Three of these judges had median correlations of .50 or below for both judgment sets.

Table 6

Frequency Distribution of Number of Pile Placement Inversions Made by Judges (280 Combinations)

Number of Inversions	Number of Judges	Judge ID[a]
0	12	
1 - 10	24	
11 - 20	10	
21 - 30	8	
31 - 40	7	
41 - 50	2	
51 - 60	2	
61 - 70	2	
71 - 80	1	
81 - 100	0	
101 - 200	4	4, 38, 64, 84
201 and Above	2	67, 83
	74	

[a] Some identification numbers for judges are higher than 74 because early lists of judges included a few company grade officers. The final set of analyses was limited to field grade officers.

Table 7

Frequency Distribution of Number of Direct Judgment Inversions Made by Judges (60 Common Combinations)

Number of Inversions	Number of Judges	Judge ID
0	23	
1 - 5	33	
6 - 10	8	
11 - 15	5	
16 - 20	3	
21 - 40	0	
41 and Above	2	4, 53
	74	

Table 8

Frequency Distribution of Median Correlations Between Judges

Median Correlation With Other Judges	Pile Placement (280 Combinations) No. of Judges	Pile Placement (280 Combinations) Judge ID	Direct Judgment (60 Common Combinations) No. of Judges	Direct Judgment (60 Common Combinations) Judge ID
0 - .10	1	56	0	
.11 - .20	2	64, 83	1	64
.21 - .30	2	38, 67	1	53
.31 - .40	0		1	67
.41 - .50	0		1	83
.51 - .60	3		4	
.61 and Above	66		66	
	74		74	

The frequency distribution of the mean pile placements did not indicate that any judges had mean pile placements that were out of line or atypical. However, for the direct judgments, the means of the values assigned the 60 common combinations by three of the judges were considerably higher and those assigned by one judge were considerably lower, than the mean values assigned by the rest of the judges. Examination of the frequency distributions of the standard deviations of the pile placement and direct judgment utilities assigned by the judges indicated that only three judges had fairly atypical indexes.

Table 9 summarizes the information presented above by showing all the judges who had atypical values for one or more of the eight indexes used. Of the 11 judges listed, six had two or more atypical values for number of inversions and/or median correlations with other judges. Using the rule of thumb adopted earlier, these six judges were removed from the sample. A seventh judge, identification number 56, was also removed because this judge had a median pile placement correlation with other judges of .05, the lowest recorded for any judge.

Before removing these seven judges, we made a check was made to see whether as a group they were in basic agreement with one another. (Here we were trying to avoid possibly eliminating a coherent minority of judges who simply had a different point of view concerning the utility of the various combinations than did the majority of field grade officers.) The intercorrelations across the 60 common combinations among the seven judges' utilities were calculated for both the pile placement and the direct judgment methods. Their median intercorrelations were .15 for the pile placements and .30 for the direct judgments (the remaining 67 judges had median

Table 9

Judges With One or More Problems in Their Judgmental Data

	No. of Inversions[b]		Median Correlations		Mean Utility		Utility SD	
Judge ID[a]	Pile PL	Direct	Pile PL	Direct	Pile PL	Direct	Pile PL	Direct
* 4	X	X				X		X
16								X
*38	X		X			X		
45						X		
*53		X		X				
*56			X					
*64	X		X	X		X		
*67	X		X	X				
*83	X		X	X				
84	X							
86								X

[a] * indicates judge's data were removed from later analyses.
[b] Pile PL = Pile Placement; Direct = Direct Judgment.

intercorrelations of .50 and above for both the pile placement and direct judgment methods). That is, they agreed neither with the other judges nor among themselves. Consequently, their judgments were excluded from the final analyses.

Imputation of Missing Data

A number of the judges had failed to provide utilities for all the MOS/performance level combinations they had been assigned. Four judges did not record their direct judgment utilities for one or two combinations. A larger number, 23, did not place one or more combinations in any of the seven utility level piles because they were not familiar enough with the job to assign a utility value. Table 10 shows a frequency distribution by number of unsorted combinations.

Although the missing judgments constituted only about 1 percent of the total data set, there was some concern that the average scale values of some MOS might be unduly affected by not being based on the same set of judges as other MOS. Consequently, the missing values were imputed, using a multiple regression procedure. Treating each judge with missing data as a dependent

Table 10
Frequency Distribution of Unsorted MOS/Performance Level Combinations (Pile Placement Method)

Number of Unsorted Combinations	Number of Judges
0	51
1 - 10	15
11 - 20	2
21 - 30	4
31 or More	2
	74

variable and judges with complete data as candidate independent variables[2], analysts used a stepwise variable selection routine to select judges whose utility ratings or pile placements added the most to the prediction of the known utility values of the judges with missing data. Judges (independent variables) were allowed to enter the multiple regression equation provided that their *F* ratio to enter was significant at the .10 level. The multiple correlation coefficients obtained for the most part were .90 or higher.

Comparison of Nonedited and Edited Data

The cumulative effect of removing seven judges and imputing pile placement or direct judgment utilities for other judges was assessed through comparing the intercorrelations and reliabilities of the nonedited and edited data sets. Table 11 shows that for the pile placement data, the 1-rater and *n*-rater reliabilities did improve for the three decks (B, D, and E) for which one or more judges had been removed. The 1-rater reliabilities improved from about .58 to .73 on the average. In contrast, the imputation of pile placement values for the combinations that were not missing one or more judges had practically no effect on the obtained reliabilities.

The 1-rater and *n*-rater reliabilities of the 60 common MOS/performance level combinations were also obtained for both the nonedited and edited ratio scale data. The 1-rater reliability rose from .564 for the nonedited data to .653 for the edited data and the n-rater reliability (based on 74 judges) increased from .990 to .992. When the pile placement reliabilities for the 60 common combinations were computed for comparison purposes, the 1-

[2]Here, the correlations are computed over the MOS/performance level combinations rated in common by the judges. The performance percentiles assigned to the combinations were used as an additional independent variable.

rater reliability rose from .673 to .746, while the n-rater reliability rose from .993 to .995. The high values for these n-rater reliabilities indicate that not much is to be gained by such editing when the number of judges is large.

Table 11

Intraclass Reliabilities for Pile Placement Data by Deck (Common and Noncommon Combinations)

	Nonedited Data			Edited Data		
Deck	No. of Raters	1-Rater	n-Rater	No. of Raters	1-Rater	n-Rater
A	12	.778	.977	12	.774	.976
B	13	.548	.939	11	.707	.964
C	12	.718	.968	12	.717	.968
D	13	.622	.954	11	.752	.971
E	12	.562	.937	9	.733	.961
F	12	.657	.958	12	.658	.958
	Average	.648	.956		.724	.966

The correlations between the utilities obtained for the nonedited and edited data were very high. Table 12 shows that even for the decks (B, D, and E) where judges were dropped because they were not typical, the correlations of mean pile placement across the 280 combinations averaged .990. The corresponding correlation for the mean direct judgment utilities assigned the 60 common combinations using nonedited and edited data was .999. While it is apparent from these results that not much was gained by editing the data, it is also evident that the editing did not unduly affect the resultant relative utility values.

Scaling Method Reliabilities

The small differences in reliability obtained for the edited vs. nonedited data should not obscure the finding that the average scale values assigned to the MOS/performance combinations had a very high degree of reliability. The n-rater reliabilities for the six separate decks ranged from .958 to .976 for the edited pile placement data. The n-rater (67 judges) reliability for the edited direct judgment utilities of the common combinations was .992. The corresponding reliability for the pile placements of the common combinations (across all decks and the 67 judges) was .995. The correlation obtained between the average scale values from the two methods across the 60 common combinations was .98.

Table 12

Correlation Between Mean Pile Placement Using Nonedited and Edited Data

Deck	Number of Combinations	Edited vs. Nonedited Correlation
A	280	.9998
B	280	.9930
C[a]	270	.9999
D	280	.9929
E[b]	275	.9856
F	280	.9998

[a] MOS 16L and 27Q were rescinded as of 31 October 87, so their data were deleted.
[b] MOS 24W was rescinded as of 31 October 1986, therefore its data were deleted.

This high correlation is not wholly attributable to judges simply assigning higher values to combinations with higher percentiles. This can be seen by the correlations between average pile placement and direct judgment utilities attained when the correlations are computed across the 12 common MOS holding percentile level constant. These correlations, presented in Table 13, had an average value of .77. The n-rater (67 judges) reliabilities averaged .89 and .82 respectively for the pile placement and direct judgment utilities when the reliabilities were computed for each percentile level separately.

These results clearly demonstrate that the judges were reacting to more than the percentile levels assigned to the combinations in making their utility judgments. Figure 4 shows the bivariate plot of mean pile placement and mean direct judgment by percentile level for the 60 common combinations.

Illustrative Interval Scale Utilities

It is of interest to note which MOS/performance level combinations received the lowest and highest utilities. Does the pattern seem reasonable? Do the relative values make sense given what we know about the MOS? Table 14 presents the noncommon combinations that received mean pile placements of 1.1 or less by the 9 to 12 officers who judged them. (The lowest or negative utility pile was assigned a value of 1.0.) Poor performance (10th percentile) was assigned the most negative utility for repairers of complex military equipment and for the Cardiac Specialist and Pharmacy Specialist. In each of these MOS, the consequences of poor performance are judged to be especially costly.

Table 13

Correlations and Reliabilities of Pile Placement and Direct Judgment Scale Utilities Holding Percentile Level Constant

Percentile Level	Number of Common Combinations	Correlation of Mean PP with DJ[a]	n-Rater Reliability Pile Placement	n-Rater Reliability Direct Judgment
10	12	.85	.89	.67
30	12	.84	.90	.68
50	12	.59	.88	.83
70	12	.63	.82	.95
90	12	.95	.94	.97
Average		.77	.89	.82

[a] PP = Pile Placement; DJ = Direct Judgment

Table 15 lists the noncommon combinations that received the highest mean pile placements (means of 6.9 or more on a scale of 7). This list includes both repairers and operators of advanced weapon systems. High performance in two intelligence MOS, 96B and 98C, was also assigned a high utility. In comparison, the mean pile placement for the 90th percentile Infantryman (11B) was 6.6.

Comparison of Utility Ratings by Different Officer Groups

The field grade officers who served as judges came from a variety of Army backgrounds. Analyses were conducted to determine whether officers in different primary specialties assigned significantly different utilities to the common MOS/performance level combinations. First, the officers in the Combat branches were placed into four categories -- Armor, Aviation, Infantry, and Other Combat (Air Defense and Artillery). Then, the separate means of the direct judgment utilities of each of the 59 common combinations were compared using analysis of variance.

To further test whether the type of judge influenced utility ratings, a separate linear regression equation was computed for each of the 12 common MOS for each judge, using performance percentiles as the independent variable and the direct judgment utilities assigned by the officer as the dependent variable. The mean slopes and the mean y-intercepts or equation constants of the officer groups were then analyzed for significant differences, using separate analysis of variance for the 12 regression slope means and the 12 intercept means.

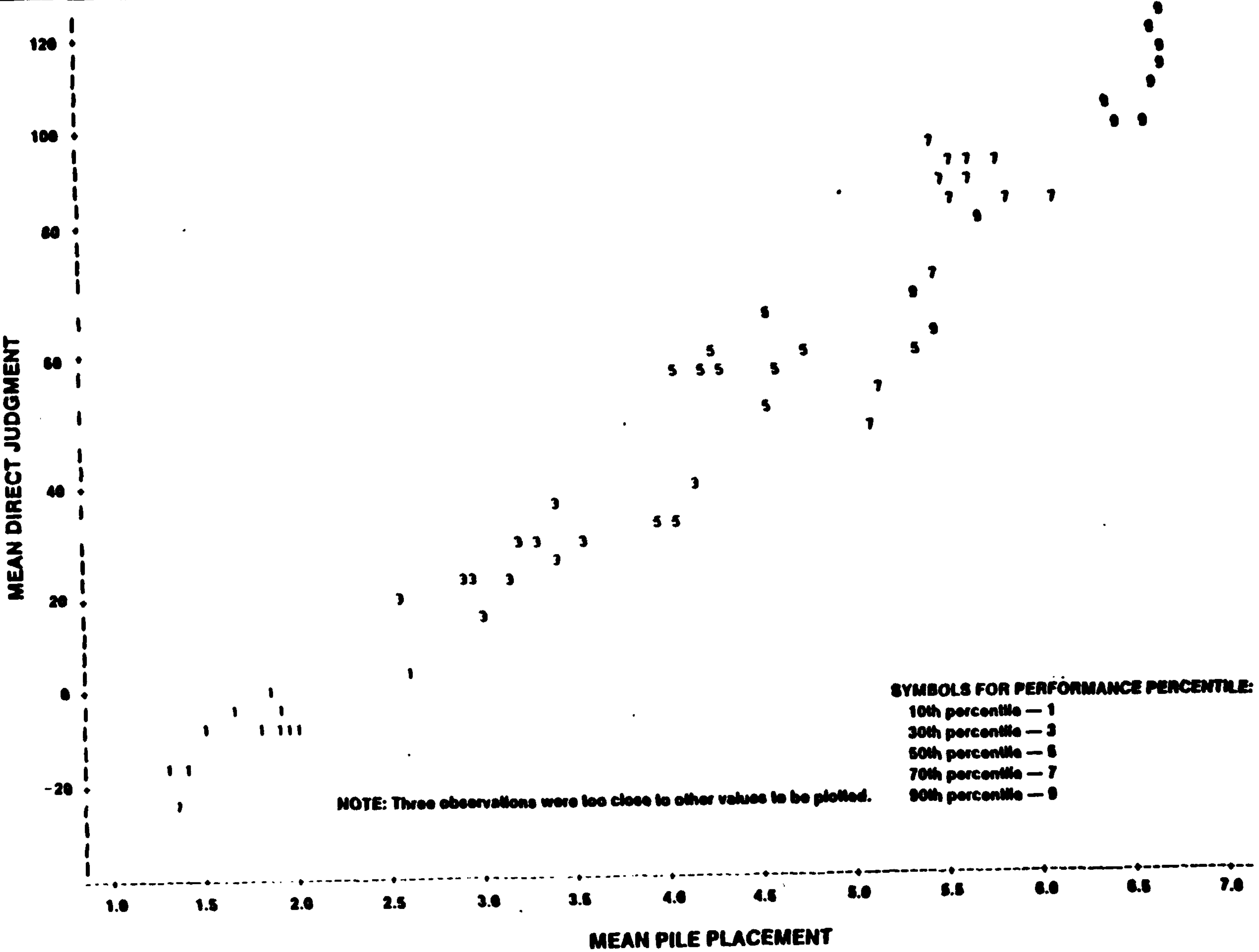


Figure 4. Plot of mean pile placement and direct judgment utilities: Sixty common combinations (67 judges)

Table 14

List of MOS/Performance Level Combinations Receiving Lowest Pile Placements (Mean Placement 1.1 and Below)

MOS ID[a]	MOS Name
23N	NIKE/HERCULES Track Radar Repairer
34Y	Field Artillery Tactical Fire Direction Systems Repairer
35R	Avionic Special Equipment Repairer
63N	M60A1/A3 Tank System Mechanic
67N	Utility Helicopter Repairer
67R	AH-64 Attack Helicopter Repairer
67Y	AH-1 Attack Helicopter Repairer
68B	Aircraft Powerplant Repairer
68F	Aircraft Electrician
91N	Cardiac Specialist
91Q	Pharmacy Specialist

[a]All associated performance levels were at the 10th percentile.

Table 15

List of MOS/Performance Level Combinations Receiving Highest Pile Placements (Mean Placement 6.9 and Above)

MOS ID[a]	MOS Name
11H	Heavy Anti-Armor Weapons Infantryman
13C	TACFIRE Operations Specialist
13R	Field Artillery Firefinder Radar Operator
15E	Pershing Missile Crewmember
15J	MLRS/Lance Operations/Fire Directions Specialist
16J	Defense Acquisition Radar Operator
21L	Pershing Electronics Repairer
24E	Improved Hawk Fire Control Mechanic
24L	Improved Hawk Launcher and Mechanical Systems Repairer
24P	Defense Acquisition Radar Mechanic
34L	Field Artillery Digital Systems Repairer
35R	Avionic Special Equipment Repairer
45N	M60A1/A3 Tank Turret Mechanic
45T	Bradley Fighting Vehicle System Turret Mechanic
63B	Light Wheel Vehicle Mechanic
63N	M60A1/A3 Tank System Mechanic
67R	AH-64 Attack Helicopter Repairer
67X	Heavy Lift Helicopter Repairer
68B	Aircraft Powerplant Repairer
96B	Intelligence Analyst
98C	EW/SIGINT Analyst

[a]All associated performance levels were at the 90th percentile.

Similar sets of analysis of variance tests were run comparing the utilities assigned by Combat, Combat Support, and Combat Service Support officers. In addition, the utilities assigned by the 47 majors in the sample were compared to those assigned by a combined group of 20 lieutenant colonels and colonels.

In all, only 10 of the more than 250 statistical tests run were significant at the .05 level. Examination of the significant differences that were obtained did not reveal any trend in the data indicating that certain types of officers favored particular MOS or performance levels.

Estimation of Ratio Scale Utilities From Pile Placement (Interval) Data

A basic objective of the overall research design was to place all 1380 MOS/performance level combinations on the same utility scale. Using the averages (across all judges) of the direct judgment utilities assigned the 60 common combinations as the dependent variable, and the pile placement of the same common combinations as the basic independent variable, an equation was derived for each separate group of judges expressing direct judgment utilities as a function of average pile placement.

This equation was then used to estimate the ratio scale values (direct judgment utilities) that each group would have assigned all the noncommon MOS/performance level combinations if they had been given that scaling task. It was assumed that since these equations would place all the estimated utility values on the same scale, minor group differences in pile placements of the 60 common combinations would be averaged out. (As the judges were assigned randomly to decks, any differences among the groups in mean pile placements could be attributed to sampling error.[3])

To explore the use of alternative regression equations for estimating the ratio scale values from the pile placement data, a subset of 20 of the common combinations were temporarily set aside and not used in the initial derivation of the reregression equations. These 20 combinations came from four MOS having performance percentiles that had utilities fairly evenly spread across the range of ratio scale utilities (based on the 67-judge averages).

Since both estimated and actual values would be available for these 20 combinations, the ability of alternative regression equations to estimate ratio scale utilities from the pile placement data could be evaluated or cross validated. That is, a regression equation from each deck based on 40 common combinations could be used to estimate the ratio scale values of the 20 set-aside combinations for which actual values were available.

---

[3]Analysis of variance significance tests run on the 59 common combinations to compare thei. mean pile placements by deck resulted in only one significant difference, a result easily attributable to chance.

Two indexes of how well the estimated ratio scale values corresponded to the actual values were (a) the mean difference between the actual and the estimated values, and (b) the square root of the mean square of the difference between the actual and estimated values. The two indexes were computed using five different sets of independent variables:

1) Average pile placement of the 40 combinations.

2) Average pile placement and the performance percentiles of the 40 combinations.

3) Average pile placement and the square of the average pile placement.

4) Average pile placement and the cube of the average pile placement.

5) Average pile placement and both the square and cube of average pile placement.

The square and the cube of the average pile placements were used as independent variables because the by-deck bivariate plots of the average ratio scales (the dependent variable) versus the average pile placement (the independent variable) suggested that there might be inflections in the best fitting lines at the two ends of the utility continuum. This might be brought about by the restriction inherent in the procedure used that limited the judges to seven utility levels when placing the combinations into piles.

Table 16 shows the results of these analyses. The equations based on all five sets of independent variables tended to over-estimate somewhat the actual utilities of the 20 holdout combinations. This tendency was most pronounced for combinations having intermediate actual utilities. In general, the equations tended to underestimate the utilities of the holdout combinations having high actual utilities, and slightly overestimate the utilities of the combinations having low actual utilities. The best balance was achieved by using average pile placement and both its square and cube as the independent variables.

The lowest mean squares for prediction errors was also obtained by the equations that used the average pile placement and both the square and the cube of average pile placement. For all equations, the largest squared errors were for the combinations having actual abilities in the mid-range. As mentioned above, the equations slightly, but consistently, overestimated these utilities.

Based on these data, pile placement and its square and cube were used as the independent variables. The equations for each deck were recomputed, using the pile placement and direct judgment values for all 60 combinations. Table 17 presents the adjusted correlation coefficients (R^2) obtained for each deck, as well as the actual equation weights. The multiple correlation coefficients remained high (about .97 on the average). The sign of the weights obtained formed a fairly consistent pattern with average pile

Table 16

Comparison of Equations Estimating Ratio Scale Utilities From Pile Placement Data (n = 20 Combinations in Each of 6 Decks)

Equation Independent Variables	Average $(A - E)^a$	Sq. Root of Average $(A - E)^2$	Average R^2 of Decks[b]
Average pile placement (PP)	- .87	9.15	.943
PP and performance percentile	-1.65	10.01	.950
PP and $(PP)^2$	- .86	9.04	.944
PP and $(PP)^3$	- .86	9.01	.944
PP, $(PP)^2$, and $(PP)^3$	- .79	8.96	.945

[a]Actual ratio scale utility (A) minus estimated ratio scale utility (E).
[b]Forty common combinations were used in each deck to obtain the squares of the multiple correlation coefficients ($\underline{R}^2$).

Table 17

Multiple Regression Equations for Estimating Ratio Scale Utilities From Average Pile Placement Data (Based on 60 Common Combinations)

Independent Variable	Deck A	Deck B	Deck C	Deck D	Deck E	Deck F
Pile Placement (PP)	14.00	21.81	43.39	24.09	46.99	49.45
$(PP)^2$	1.455	-.323	-5.785	-1.344	-6.922	-6.932
$(PP)^3$	-.053	.067	.485	.169	.545	.549
Intercept	-34.09	-41.75	-69.44	-47.74	-63.80	-77.85
Adjusted $\underline{R}^2$	.965	.926	.954	.944	.912	.924

placement always having a positive weight, and, with the exception of Deck A, the square and cube of average pile placement having negative and positive weights, respectively.

Cross-Validation of Estimation Equations on a Holdout Sample

How would the utility values derived from the equations given in Table 17 compare with utilities obtained from direct judgments by field grade officers? To explore this question, the participants[4] in the last utility workshop were given an additional 40 combinations (8 MOS x 5 levels) on which to make their direct judgments of utility. The mean of the direct judgment utilities given these 40 combinations by these officers could then be compared to values from the other 58 officers computed by formulas derived for each deck excluding the data obtained from the last workshop.

Before comparing the utilities, the set of 40 new utility estimates from the holdout sample were adjusted so that the new values corresponded to the utilities that could have been expected if the remaining sample of 58 judges had actually evaluated the additional 40 combinations using the direct judgment procedure.

A multiple regression equation was derived using the mean direct judgment utility assigned the 60 common combinations by the 58 officers as the dependent variable, and the mean utilities assigned the 60 combinations by the nine officers, the square of these utilities, and the performance percentile of the 60 combinations as the independent variables. The estimated direct judgment values for the 40 extra combinations were then obtained using this equation. This procedure adjusted the holdout sample utilities for random and nonrandom differences between the holdout and remaining sample judges in the direct judgment of the 60 common combinations.

As shown in Table 18, very high correlations (.97) were obtained between the utilities estimated from the separate deck equations and the holdout sample unadjusted and adjusted direct judgment utilities. Moreover, the utility means, standard deviations, and ranges for the 40 extra combinations obtained from the holdout sample direct judgments were quite similar to those estimated from the deck equations. This was especially evident after adjustment for utility differences between the two officer samples on their direct judgment of the 60 common combinations.

---

[4]None of the 10 participants were Armor officers, unlike the remaining sample where 17 of the 58 officers were Armor officers. One of the 10 was Judge ID #83 who, as noted earlier, was removed from the sample because of a large number of inversions and low medium correlations with the other judges. Hence final analyses from this workshop were based on nine participants.

Table 18

Comparison of Utilities Obtained for 40 Extra MOS/Performance Level Combinations

Origin of Utility Estimate	Judges	Mean	Standard Deviation	Range
(1) Direct judgments by holdout sample of officers	9	62.8	38.4	-9.7 to 123.7
(2) Variable 1 adjusted for differences between the hold-out and remaining samples on 60 common combinations	9,58	54.2	36.5	-5.8 to 110.8
(3) Separate deck equations based on data from remaining officer sample only	58	56.3	36.1	-8.2 to 111.7

Intercorrelations

	(1)	(2)
(1)	---	---
(2)	.993[a]	---
(3)	.972	.973

[a]Variable 1 was used in the derivation of Variable 2.

THE FINAL UTILITY VALUES

All the analyses described were performed for the purpose of establishing the reliability and validity of the utility estimates for MOS x performance level combinations obtained by using the technique described in this report.

The analyses support the conclusions that:

1) For both methods the reliability of a single judge is reasonably high.

2) For both methods the reliability of the average value produced by 11 judges or more is _very_ high.

3) Reliabilities are high even when performance level is controlled and differences are due only to MOS differences within performance level.

4) The agreement between the two methods when scale values are compared on a set of common stimuli is very high and equal to the limit of their reliabilities.

5) Officers of different ranks or MOS specialties do not produce different patterns of scale values.

6) Patterns of high and low utility values "make sense".

7) A relatively simple exercise in equation fitting produced a useful method for estimating ratio scale values (which could not be obtained for all MOS x performance level combinations) from the interval scale values obtained from all MOS x performance level combinations using the pile placement method.

8) As determined on a cross-validation sample of stimuli, the equations used to estimate ratio values from interval data were highly accurate ($R_{estimated}$ x actual = .97).

All of these reliability and validity analyses were based on data that had been carefully edited for missing data and outliers.

<u>Utilities for MOS/Performance Level Combinations</u>

The derived equations for each deck were used to estimate the ratio scale utilities for the noncommon MOS/performance level combinations for the entry-level list of 273 MOS. These values are given in Appendix B, along with the actual average direct judgment utilities for the 60 common combinations.

Consequently, Appendix B represents the "bottom line" of the Project A utility scaling work to date. It contains ratio scale utility values for 273 x 5, or 1365, MOS/performance level combinations. Within the limits of the reliability and validity evidence discussed in this report, the 1365 combinations have been placed on the same scale. As an example of this extended MOS list of utility values, the ratio scale values for the 19 Project A MOS are shown in Table 19.

<u>Equations for Estimating Utilities for Continuous Performance Distributions</u>

To make it possible to assign a utility value to any performance percentile within an MOS (not just the 10, 30, 50, 70, and 90 percentiles), a separate equation was derived for each of the 273 MOS relating performance percentile and the square and cube of performance percentile (the independent variables) to utility (the dependent variable).

Table 19

Average Ratio Scale Utilities by Performance Percentile For the Project A MOS

MOS	Percentile 10	30	50	70	90
11B*	4.5	38.7	61.7	82.3	100.0
12B	5.1	47.2	63.3	86.8	112.6
13B	20.8	35.0	67.1	85.2	111.7
16S	-8.2	33.4	40.3	77.3	105.6
19E	8.4	45.6	68.3	98.0	108.2
27E	-13.8	31.0	50.7	91.1	115.2
31C	-6.4	25.4	54.3	86.8	97.9
51B	22.7	36.9	54.0	65.9	59.3
54E	1.8	31.7	45.4	89.1	108.7
55B	0.6	40.1	61.5	82.6	100.2
63B	0.5	35.0	55.1	87.7	111.7
64C	22.7	48.8	52.2	81.9	79.6
67N	-24.0	8.5	46.9	80.5	107.7
71L	0.5	27.3	54.3	70.8	86.8
76W	15.5	39.2	59.1	73.5	82.7
76Y	-5.0	33.1	66.4	81.5	92.7
91A	-4.0	17.6	52.5	76.6	100.2
94B	2.7	27.4	63.9	85.5	90.8
95B	-8.2	38.6	63.1	84.2	108.7

*One of 12 common MOS assigned actual ratio scales

Each equation was based on five data points, the estimated (or actual) average ratio scale utilities respectively assigned the 10th, 30th, 50th, 70th, and 90th percentile levels within the MOS. To determine the general shape of the relationship between percentile level and utility across the MOS, a stepwise multiple regression procedure (SAS) was used. The order of entry of the independent variables into the equation and the significant levels and signs (positive or negative) of their regression weights were noted for each MOS equation.

Table 20 summarizes the equations derived for the 273 MOS. Performance percentile was the first independent variable to enter into the equation for each MOS. For 91 of the MOS equations, a second independent variable entered the equation with a statistically significant weight (.10 level of significance). For the most part (95%), this second variable entered with a negative weight, indicating that the rate of increase in utility was declining for higher performance percentiles. The cube of performance

Table 20

Summary of MOS-Specific Equations Derived for Estimating the Utility of Performance at Different Percentile Levels

	First Variable Entering Significantly Into Equation		Second Variable Also Entering Significantly Into Equation		Third Variable Also Entering Significantly Into Equation	
	Sign Of Regression Weight		Sign Of Regression Weight		Sign Of Regression Weight	
Independent Variable	+	-	+	-	+	-
Performance percentile (P)	273	0	0	0	0	0
Percentile squared (P^2)	0	0	1	38	1	1
Percentile cubed (P^3)	0	0	4	48	1	0
Total	273	0	5	86	2	1

percentile was selected as the second equation variable a little more often than the square of performance percentile (52 vs. 39 times). For only three of the 91 equations did a third independent variable also enter the equation significantly (.10 level). This would indicate a best fitting line with two inflection points.

Figure 5 shows bivariate plots between utility and performance percentile within three MOS. Line B in the figure shows a typical plot when the relationships are essentially linear between utility and performance. An example of a plot where a second variable (percentile squared) enters into the regression equation with a negative weight is provided by line A while line C shows an example where the second independent variable entering the equation is percentile cubed.

Notice the very high values of adjusted R^2 in the examples shown. The five data points in the plots were determined from utility judgment data. As shown earlier (see Table 13), reliable differences between the utilities assigned MOS are obtained at all percentile levels when performance percentile is held constant. Within MOS, however, utility is highly predictable from performance percentile, though the relationship is frequently not linear over the range of performance.

The operational significance of these findings is that the utility of assigning a recruit to any MOS can be estimated using a within-MOS equation to relate the level of the recruit's predicted performance in the MOS to the recruit's utility for that MOS. These utilities, in turn, could be used to help decide the MOS to which the recruit should be assigned under an algorithm for optimizing job assignments.

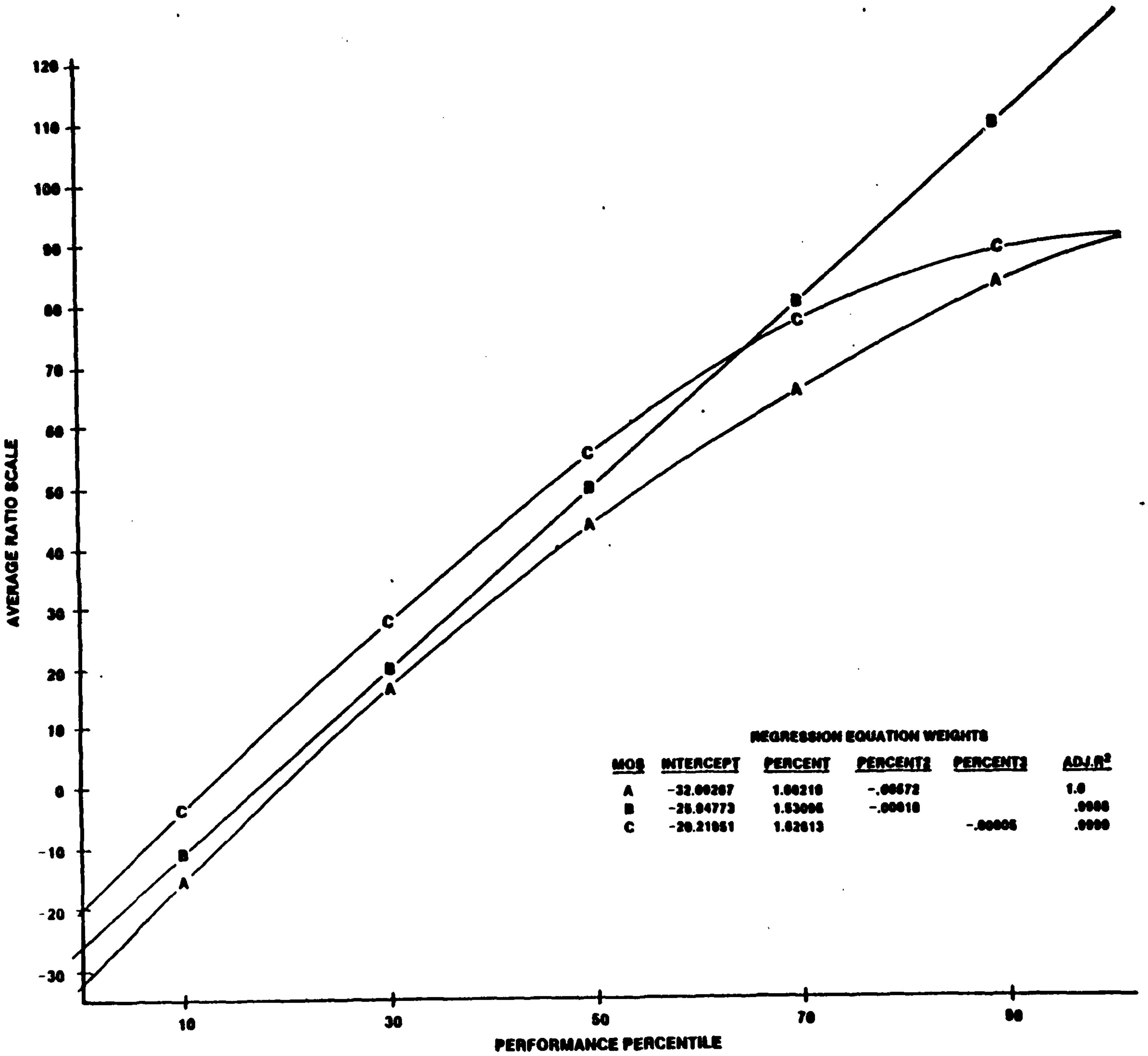


Figure 5. Average ratio scales by performance percentile for three MOS

DISCUSSION AND CONCLUSIONS

Perhaps the most important finding in this first research effort to establish utility values for different levels of performance in all Army entry-level MOS is that it can be done. One can fault the subjective nature of the judgments called for, but the fact remains that the mean utilities had very high reliabilities, both within and between methods (for example, for the common set of combinations, .99 within the utility scaling method and .98 between methods). The high correlation (.97) and similarity between the utilities derived from separate deck equations and the utilities assigned directly by the holdout officer sample (Table 18) further attests to the stability of the utilities across methods and officer samples. Similar results have been reported by Bobko and Donnelly (1988), who were seeking to identify elements that enter into job-level estimates of the overall value of job performance. They found that such subjective ratings of job performance value could be predicted and modeled to a substantial degree (both within and across performance levels) from a variety of job-level predictors, in a variety of content domains.

The high reliabilities of the mean utilities assigned the MOS/performance level combinations, and the lack of any clear pattern of differences in average utilities assigned by officers from different MOS specialties, also indicate that similar values would result if the utilities were assigned by a different sample of officers than the one used in this research. Field grade Army officers apparently share similar perceptions of the relative worth/costs of low and high performance in Army MOS. This research clearly demonstrates that this shared organizational value function can be reliably scaled.

The finding that consistent differences in the utility of performance are obtained for different MOS at each percentile level is also worth noting, especially when combined with the finding that the relative ranking of the MOS in terms of their utility levels shifts depending upon the predicted performance level. A personnel assignment algorithm that took into account these utility differentials at all performance levels would most likely be able to effect more optimal Army-wide assignments than one that did not (Nord & White, 1988).

However, a number of critical problems need to be addressed before utilities similar to the ones obtained in this research can be used operationally. Foremost among these perhaps is the problem of how to ensure that the proper distribution of available personnel talent is assigned to each MOS. An assignment algorithm that paid attention only to the utility of assigning individual recruits to MOS or CMF, without regard to the utility of the total distribution of low- or high-quality personnel being assigned to each job, could result in certain jobs being filled by insufficient numbers of technically proficient recruits. Research to determine the utility of different distributions of available recruits in Army jobs is the subject of an ongoing parallel effort being undertaken by ARI researchers (Nord & White, 1987).

Another issue concerns obtaining the acceptance of these utility values by those who are responsible for personnel policies and decisions. Such approval is unlikely to occur unless it can be demonstrated that use of the utility information would result in more optimal manpower allocations. Work is now in progress (Nord & White, 1987) to examine the effects of using (or not using) these job-specific utility functions to make personnel classification and job assignment decisions.

Yet another issue concerns the duration of time that the recruits actually remain in the Army. All other things being equal, recruits who complete their first tour of enlistment will have higher utility than those who do not. Similarly, high-quality soldiers who reenlist will have more utility to the Army than those who fail to reenlist.

A related consideration is cost. Recruiting, training, maintaining, and retaining high-quality soldiers is a costly operation. These costs are not equal across Army jobs. It costs more, for example, to recruit high-quality personnel than it does to obtain recruits of lower quality. Likewise, it may take longer and be more costly to train soldiers in high-technology career management fields than in some other types of CMF. Potential cost, reenlistment propensity, and NCO potential all should obviously be considered, if at all possible, in making the initial assignments of recruits to jobs.

Finally, assuming that judged utility of performance has a role in an optimal classification and job assignment system, questions remain concerning how the requisite judgments should be obtained operationally. What types of officers should make the judgments involved? How often do the resultant utility functions have to be updated to keep the utilities current? Do the utilities of all entry-level MOS have to be determined, or can the utilities for most MOS be inferred from those assigned a representative sample of MOS taken from career management fields or other MOS groupings?

Clearly, this research has affirmatively answered the question of whether a coherent, reliable set of relative utility values could be derived for all performance levels in all entry-level Army MOS. The next steps involve how to make best use of that finding in improving the Army's selection, classification, and assignment processes.

BLANK PAGE

REFERENCES

Bobko, P., & Donnelly, L. (1988). Identifying correlates of job-level, overall worth estimates: Application in a public sector organization. Human Performance, 1(3), 187-204.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 37, 65-76.

Brogden, H. E. (1949). When testing pays off. Personnel Psychology, 2, 171-183.

Campbell, J. P. (Ed.). (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). (AD A193 343)

Campbell, J. P. (Ed.). (1986). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). (AD A198 856)

Campbell, J. P. (Ed.). (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1987 fiscal year (ARI Technical Report in preparation).

Cascio, W. F. (1982a). Applied psychology in personnel management (2nd ed.). Reston, VA: Reston Publishing.

Cascio, W. F. (1982b). Costing human resources: The financial impact of behavior in organizations. Boston: Kent Publishing.

Cronbach, L. J., & Gleser, G. C. (1965). Psychological tests and personnel decisions. (Rev. ed.) Urbana: University of Illinois Press.

Eaton, N. K., & Goer, M. H. (Eds.). (1983). Improving the selection, classification, and utilization of Army enlisted personnel: Technical appendix to the annual report (ARI Research Note 83-37). (AD A137 117)

Eaton, N. K., Goer, M. H., Harris, J. H., & Zook, L. M. (Eds.). (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year (ARI Technical Report 660). (AD A178 944)

Hoffman, R. G. (1987). Clustering Army Military Occupational Specialties for Project A: Phase Two (HumRRO IR-PRD-82-22). (ARI Technical Report in preparation.)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1983). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report (ARI Research Report 1347). (AD A141 807)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report synopsis, 1984 fiscal year (ARI Research Report 1393). (AD A173 824)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Appendices to annual report, 1984 fiscal year (ARI Research Note 85-14).

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1985). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year - Supplement to ARI Technical Report 746 (ARI Research Note 87-54).

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1986). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year - Supplement to ARI Technical Report 792 (ARI Research Note 88-36).

Nord, R., & White, L. A. (1987). Optimal job assignment and the utility of performance. Paper presented at the Annual Meeting of the American Psychological Association, New York.

Nord, R., & White, L. A. (1988). The measurement and application of performance utility. In B. F. Green, H. Wing, & A. K. Wigdor (Eds.), Linking military enlistment standards to job performance. Washington, DC: National Academy Press.

Roulon, P. J., Tiedman, D. V., Tatsuoka, M. M., & Langmuir, C. R. (1967). Multivariate statistics for personnel classification. New York: John Wiley & Sons.

Sadacca, R., & Campbell, J. P. (1985). Assessing the utility of a personnel/classification system. Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, Georgia. (In ARI Research Note 87-54). (AD A188 267)

Sadacca, R., Campbell, J. P., White, L. A., & DiFazio, A. S. (1988). Weighting criterion components to develop composite measures of job performance (ARI Technical Report 838).

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work force productivity. Journal of Applied Psychology, 64, 609-626.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussions and tables. Journal of Applied Psychology, 23, 565-578.

Torgerson, W. S. (1958). Theory and methods of scaling. New York: John Wiley & Sons.

Ward, J. H., Haney, D. L., Hendrix, W. H., & Pina, M. (1978). Assignment procedure in the Air Force procurement management information system (AFHRL-TR-78-30). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Appendix A

Utility Workshop Instructions and Materials

AGENDA

Workshop on Utility Measurement

0800 - 0815	Project A Overview
0815 - 0820	Completion of Background Information Sheet
0820 - 1000	Sorting hypothetical recruits on basis of their utility
1000 - 1015	Break
1015 - 1100	Comparing hypothetical recruits to a standard
1100 - 1145	Assigning groups of recruits to MOS
1145 - 1200	Discussion

Utility Workshop Participant

Background Information Sheet

Judge No. ________

Name and Grade: ____

Primary Specialty: ____

Secondary Specialty: ____

Title of Current Position: ____

Title of Last Position: ____

Years of Service: __________ Regular Army: Yes____ No____

Types of enlisted MOS which most experience has been with: ______________

____

____

____

Assigning Utility Values to Army Recruits

Overview

Many people have suggested that, all other things being equal, newly accessioned recruits differ in terms of their predicted overall value of utility to the Army for accomplishment of particular Army missions. This is not an argument that certain recruits are not needed. Rather, it is an assertion that, given some base line, adding certain recruits to some MOS has relatively greater utility to the Army than adding other recruits to other MOS.

In order to allow a computerized enlisted personnel selection and classification system to operate in the best interest of the Army, the decisions made by the system must reflect the best judgment of experienced Army officers. To inform the computerized processes involved in selecting and classifying applicants for enlistment, you will be asked to judge the relative priority that the system should place on filling different MOS with recruits having different predicted performance levels. This is not to imply that some MOS should have no newly assigned recruits or only low or high level recruits, but that the system should attempt to meet the most critical Army personnel needs first.

Purpose of Workshop

In order to determine how best to measure MOS personnel classification priorities, we are trying out various methods of obtaining judgments of experienced Army officers. In this workshop we will try out three methods. The methods call for increasingly complex judgments concerning the value or usefulness of classifying recruits into different MOS. In the first procedure, you will be asked to sort recruits into utility categories based on relative classification priorities. In the second procedure, you will judge the value of recruits relative to the value of a specified Infantryman (11B) recruit. The third procedure involves classifying groups of recruits of various predicted performance levels into a limited number of MOS. In making all the judgments called for, please consider the likely usefulness of the recruits at the given performance level in that MOS in helping achieve the Army's mission in comparison to other recruits at other performance levels in other MOS.

Assumptions

To help assure that all workshop participants are starting from the same place, we would like you to make the following assumptions when making your judgments:

(1) The military context for which the utility of the recruits is being considered is as follows:

The world is in a period of heightened tensions. There is an increasing probability that hostilities will break out in Europe, Asia, the Caribbean, Latin America and Africa. The Army's mission is to support U.S. treaty obligations and to help defend the borders of allied and friendly nations. Some of the potential enemies have nuclear and chemical capability. Air parity does exist between allied forces and potential hostile nations. U.S. Army training and other preparatory activities have been substantially increased. Most combat and associated support units are participating in frequent field exercises. Most units are being actively resupplied.

(2) The field strength of all MOS overseas is the same--70 percent.

(3) Troop replacement needs resulting from any anticipated wartime casualties will be handled separately by the computerized personnel selection and classification system. (That is, the relative priorities should reflect only the likely usefulness of recruits at given predicted performance levels in helping to achieve the Army's mission.)

(4) The measures used to predict performance include not only aptitude scores (taken from the Armed Services Vocational Aptitude Battery) but also tests of psychomotor skills, work history, interests, motivation, and other indexes that predict overall MOS performance.

(5) The overall MOS performance measure for each MOS represents an optimally weighted (for that MOS) combination of several performance factors. Thus, recruits at the highest predicted performance level (90th percentile) in each MOS are more likely to be dependable, proficient in MOS tasks, know the facts and procedures required to do their jobs, perform more effectively under adverse or difficult conditions, avoid disciplinary problems, provide support to fellow soldiers, and be more physically fit.

(6) The predicted performance levels for the recruits are accurate. That is, the recruits will actually perform at the predicted levels.

(7) The spread or amount of variation in predicted performance is equal in each MOS.

DIRECTIONS FOR PROCEDURE A: SORTING ARMY RECRUITS ON PRIORITY FOR FILLING STRENGTH REQUIREMENTS

In the first procedure to be tried out today you will be given a set of cards. On each card there is a short description of an MOS (first tour or Skill Level 10) taken from AR 611-201, which gives descriptions of all Army MOS. Also on each card is the predicted performance level in that MOS of the given recruit. The levels have been set at the 10, 30, 50, 70 or 90th percentile level and are based on the selection/classification measures available on the recruits. All percentiles are the predicted percentiles of performance of the recruits, if <u>all</u> recruits were rank ordered in terms of their predicted performance in <u>the</u> given MOS without regard to current cut-off scores. Note that a 10th percentile performance level signifies <u>low</u> performance and a 90th percentile performance level signifies <u>high</u> performance. Also please note that the percentiles refer to percentiles within all newly accessioned recruits assuming that the recruits were rank ordered in terms of their predicted performance scores for the given MOS.

The recruits on the cards have been tentatively assigned to 56 MOS without regard to current cut-off scores. The judgment task is to sort 280 cards (56 MOS X 5 predicted performance levels) into 7 piles or categories reflecting the relative utility of the recruits described on the cards. The categories are:

High positive utility would probably result if these recruits were placed in these MOS.

Between moderate and high utility would probably result if these recruits were placed in these MOS.

Moderate utility would probably result if these recruits were placed in these MOS.

Between low and moderate utility would probably result if these recruits were placed in these MOS.

Low positive utility would probably result if these recruits were placed in these MOS.

Advantages of placing these recruits in these MOS would probably be equal to the disadvantages (expected utility = 0).

Negative utility would probably result if these recruits were placed in these MOS. (Any positive contribution would probably be outweighed by problems associated with low levels of overall performance).

<u>Specific Directions:</u>

(1) Familiarize yourself with the MOS by examining the descriptions on the cards and with the above descriptions of the 7 piles. You will be provided with a set of label cards containing the pile descriptions.

(2) Then sort the cards (which are in random order) into the 7 piles. You are free to place as many cards as you like in any one pile. If you do not feel that you are familiar enough with any given MOS to make a comparative evaluation, place the cards for that MOS in an eighth, unrated pile.

(3) When you have finished your first sort, go through the piles carefully, making any changes in the sorting you feel are appropriate.

(4) When you are satisfied with your sorts, please place the appropriate label card on top of each pile and secure each pile with a rubber band. Then use a rubber band to bind the piles together with your Name Card on top.

Thank you for your cooperation.

DIRECTIONS FOR PROCEDURE B: SETTING THE UTILITY OF ARMY RECRUITS RELATIVE TO THE VALUE OF A 90TH PERCENTILE INFANTRYMAN -11B

In the second procedure you will try out today, you will be given a set of 60 cards. These cards are a subset of the cards you sorted earlier. The cards cover 12 MOS, with recruits at the 5 predicted performance levels you evaluated earlier. The judgment task is to assign a numerical utility value to each of the 60 recruits. The values that you assign will be proportionate to the value assigned to a 90th percentile Infantryman recruit. In other words, the overall worth or utility of the 90th percentile Infantryman recruit will be used as a yardstick and the worth of all other recruits will be judged in relationship to this Infantryman recruit. (This is comparable to the use of a given platinum bar as the defined length of 1 meter or 100 centimeters in the metric system.) In this case the value of the 90th percentile Infantryman will be set at 100.

In making your judgments, please make the same set of assumptions as were made in the previous procedure. Please review these assumptions before beginning the judgment task.

Specific Directions:

(1) Familiarize yourself with the 12 MOS by examining the descriptions provided.

(2) Write the value, 100, on the 90th percentile Infantryman card which is on top of the deck you just received. (The other cards in the deck are in random order.)

(3) Then take each of the other cards and assign the soldier a utility value which reflects the worth of each recruit relative to the value of the 90th percentile Infantryman recruit which has a worth of 100. You may assign higher values than 100 or even negative values to one or more recruits if you wish (see scale provided on next page). In other words, you are free to assign any number that reflects the relative worth of the recruit being evaluated. Write the values you assign directly on the cards in the lower right hand corner.

(4) When you have gone through the deck once, please arrange the cards in numerical order from lowest to highest value.

(5) Then go through the cards once more and change any assigned value that you feel is out of line with the others (with the exception of the value of 100 assigned to the 90th percentile Infantryman). Please refer to the attached Utility Rating Scale to help resolve any scaling problems.

UTILITY RATING SCALE*

150 - Recruit is worth 50% more to the Army than an Infantryman (11B) recruit at the 90th percentile level of predicted performance.

125 - Recruit is worth 25% more to the Army than an Infantryman (11B) recruit at the 90th percentile of predicted performance.

100 - Utility to the Army of an Infantryman (11B) recruit at the 90th percentile level of predicted performance in the scenario described.

75 - Utility of this recruit is 3/4 that of an Infantryman (11B) recruit at the 90th percentile level of predicted performance.

50 - Utility of this recruit is 1/2 that of an Infantryman recruit at the 90th percentile level of predicted performance.

25 - Utility of this recruit is 1/4 that of an Infantryman recruit at the 90th percentile level of predicted performance.

0 - Advantages of having this recruit in the scenario described are equal to the disadvantages.

-25 - Use of this recruit would result in a net loss to the Army equal to the gain that would result from using a recruit with a utility value of 25.

-50 - Use of this recruit would result in a net loss to the Army equal to the gain that would result from using a recruit with a utility value of 50.

* Please note that values higher than 150 and lower than -50 can be assigned. Also, any value in between the scale points given can be assigned, that is, you are <u>not</u> restricted to the values appearing on the above scale.

Appendix B

Average Ratio Scale Utilities for 273 MOS by Performance Percentile

MOS	Percentile 10	30	50	70	90
008	-13.2	22.3	52.2	74.9	88.2
01H	-19.3	22.7	33.4	77.3	94.3
03C	12.7	35.0	57.1	63.0	67.1
05D	-9.8	29.4	53.7	79.2	113.5
05H	-19.0	27.3	52.5	84.7	112.6
05K	-3.2	19.2	53.1	88.1	110.7
11B *	4.5	38.7	61.7	82.3	100.0
11C	5.0	35.0	70.6	97.4	111.8
11H	13.6	49.0	65.2	93.3	117.9
11M	8.4	41.5	74.8	90.8	103.0
12B	5.1	47.2	63.3	86.8	112.6
12C	8.5	42.9	70.8	78.0	88.1
12E	-17.5	20.8	41.3	75.7	95.9
12F	6.5	33.1	66.2	85.5	101.9
13B	20.8	35.0	67.1	85.2	111.7
13C	-17.5	32.8	53.2	80.3	115.2
13E *	-3.3	35.7	64.6	90.5	111.9
13F	15.5	35.0	59.1	85.2	105.0
13M	16.9	33.4	59.3	89.1	97.0
13R	-11.2	21.7	55.2	83.0	116.9
13T	-13.2	18.7	56.2	66.4	79.3
15D	-7.5	35.5	55.8	81.5	105.6
15E	0.6	23.5	52.5	80.6	117.9
15J	-21.8	8.5	57.3	78.0	120.1
16D	0.5	28.2	51.3	85.2	101.9
16E	-3.1	26.4	50.5	81.9	91.7
16F	-0.1	29.5	58.2	92.7	106.9
16G	-7.5	27.3	57.8	68.3	93.1
16H *	-8.4	31.0	57.3	86.3	107.8
16J	-6.3	25.8	51.3	73.5	115.2
16P	-7.2	31.2	57.3	80.5	107.7
16R	15.5	32.8	49.3	80.3	93.1
16S	-8.2	33.4	40.3	77.3	105.6
16T	-11.6	22.3	60.2	92.7	116.8

* One of 12 MOS assigned actual ratio scales

	Percentile				
MOS	10	30	50	70	90
17B	-0.6	33.4	52.2	89.1	105.6
17C	-0.6	31.7	48.8	91.7	108.7
19D	1.6	35.0	62.3	92.7	111.8
19E	8.4	45.6	68.3	98.0	108.2
19K	5.1	47.2	72.7	95.6	110.0
21G	-21.8	8.5	51.0	80.5	110.7
21L	-17.5	28.2	51.3	75.7	111.7
22L	-8.2	24.5	52.2	84.2	84.2
22N	-13.2	20.5	50.2	83.7	106.9
23N	-22.2	25.3	45.6	66.2	103.0
23U	-16.3	17.6	59.7	70.8	107.5
24C	-13.3	16.0	59.5	93.5	113.8
24E	-13.6	30.5	45.3	85.2	111.7
24G	-13.6	26.4	45.4	81.9	99.8
24H	-11.6	25.8	54.2	85.9	111.8
24J	-14.7	23.2	49.6	68.3	103.0
24K	-19.0	17.6	56.1	80.6	110.0
24L	-19.6	12.3	42.9	78.0	116.9
24M	-17.5	25.8	53.8	80.3	108.3
24N	-5.6	24.5	52.2	81.9	121.8
24P	-6.3	20.8	47.3	78.0	115.2
24Q	-16.4	22.3	52.2	77.1	106.9
24S	-9.8	21.1	45.6	62.0	88.4
24T	-19.0	17.6	50.7	78.6	112.6
24U	-17.5	23.6	59.5	75.5	107.7
25L	-10.0	24.1	56.2	85.9	109.3
26C	4.1	33.4	47.1	84.2	99.8
26E	-14.8	20.5	56.2	77.1	95.0
26F	-12.2	29.4	49.6	77.0	98.0
26H	-13.8	23.5	49.0	80.6	112.6
26K	-13.3	23.6	44.9	80.5	113.8
26Q	-9.8	25.3	45.6	77.0	108.2
26T	-9.9	15.5	51.3	67.1	80.3
26V	-13.6	20.8	48.8	77.3	111.8
26Y	-13.2	20.5	48.3	74.9	109.3

MOS	Percentile 10	30	50	70	90
27B	-8.2	26.4	50.5	77.3	102.7
27C	-13.2	20.5	58.2	83.7	104.5
27D	-17.1	21.1	43.5	62.0	93.1
27E	-13.8	31.0	50.7	91.1	115.2
27F	-21.8	17.9	46.9	80.5	110.7
27G	-17.5	30.5	45.3	80.3	101.9
27L	-10.8	31.7	45.4	81.9	115.0
27M	-10.0	15.2	56.2	85.9	111.8
27N	-14.7	27.3	47.6	68.3	98.0
27P	-14.7	16.9	43.5	66.2	83.8
29E	-11.3	27.3	52.5	86.8	105.0
29F	-21.8	14.1	55.2	90.8	104.8
29J	-9.2	25.5	53.1	83.0	90.8
29M	-9.9	23.3	53.2	78.0	108.3
29N	-6.3	28.2	47.3	75.7	101.9
29S	-5.6	24.5	54.0	89.1	97.0
31C	-6.4	25.4	54.3	86.8	97.9
31K	-3.2	27.4	57.3	85.5	96.2
31M	-6.3	32.8	51.3	82.7	93.1
31N	-14.8	18.7	52.2	77.1	102.1
31V	-5.1	25.3	55.8	88.4	105.6
32D	-6.4	19.6	54.3	76.6	93.3
33P	-11.6	18.7	50.2	83.7	116.8
33Q	-17.1	10.5	45.6	79.2	108.2
33R	-24.4	21.6	52.5	84.7	105.0
33T	-11.2	21.7	48.9	80.5	110.7
34L	-15.4	19.8	53.1	78.0	116.9
34T	-13.6	28.2	49.3	80.3	108.3
34Y	-21.5	23.3	45.3	80.3	105.0
35E	-5.6	33.4	47.1	86.6	105.6
35G	-5.6	20.8	47.1	75.1	99.8
35H	-6.7	20.5	48.3	70.6	85.9
35K	-19.6	19.0	47.6	79.2	98.0
35L	-21.6	17.6	56.1	84.7	107.5
35M	-21.8	23.6	46.9	70.8	110.7
35R	-21.5	20.8	45.3	75.7	111.7
36C	4.6	40.9	48.9	75.5	63.0
36L	-8.2	28.2	61.2	66.9	99.8
36M	6.4	31.7	55.7	75.1	97.0
39B	-11.6	24.1	48.3	72.8	99.7

MOS	Percentile 10	30	50	70	90
41B	-14.8	22.3	44.4	72.8	81.5
41C	-19.3	28.2	52.2	81.9	105.6
41E	6.2	31.4	47.6	68.3	72.6
41J	0.6	27.3	50.7	68.9	76.6
42C	-5.6	26.4	47.1	63.1	86.6
42D	-13.2	22.3	48.3	74.9	83.7
42E	-2.8	25.3	45.6	68.3	83.8
43E	-19.6	14.8	43.5	72.6	110.8
43M	6.2	27.3	64.1	62.0	72.6
44B	0.8	35.0	57.3	70.8	90.8
44E	-9.9	35.0	51.3	80.3	98.9
45B	6.8	39.2	51.3	85.2	95.9
45D	6.4	33.4	55.7	75.1	97.0
45E	-13.2	29.5	60.2	90.4	114.3
45G	-11.6	24.1	58.2	90.4	116.8
45K	-7.5	27.3	59.9	79.2	108.2
45L	-1.7	31.0	59.7	86.8	115.2
45N	-7.2	31.2	63.9	83.0	116.9
45T	-13.6	25.8	47.3	85.2	115.2
46N *	-15.0	25.9	57.5	93.1	119.5
51B	22.7	36.9	54.0	66.9	59.3
51C	-8.3	20.5	46.4	72.8	83.7
51G	3.3	31.3	52.2	66.4	74.9
51K	10.5	25.3	59.9	66.2	70.5
51M	10.5	33.4	53.7	74.8	79.2
51N *	-7.6	29.3	51.5	69.6	81.5
51R	-1.7	19.6	47.2	67.0	84.7
52C	1.8	31.7	59.3	66.9	91.7
52D	-3.4	24.1	60.2	81.5	92.7
52F	-14.7	27.3	47.6	66.2	95.6
52G	-11.2	25.5	53.1	73.1	85.5
54C	9.8	35.0	55.1	80.3	87.7
54E	1.8	31.7	45.4	89.1	108.7
55B	0.6	40.1	61.5	82.6	100.2
55D	-13.2	20.5	54.2	70.6	97.4
55G *	-24.0	19.7	56.4	94.1	124.8
55R	-1.2	33.1	59.5	73.1	90.8

* One of 12 MOS assigned actual ratio scales

MOS	10	30	50	70	90
	Percentile				
57E	23.5	25.4	52.5	68.9	72.7
57F	2.7	35.0	63.9	75.5	75.5
57H	12.7	41.3	53.2	85.2	90.4
61B	8.6	38.6	42.0	75.1	89.1
61C	-1.7	17.6	49.0	65.2	88.9
62B	0.8	31.2	66.2	85.5	96.2
62E	-6.7	31.3	60.2	85.9	92.7
62F	-0.5	37.5	57.8	83.8	86.1
62G	0.6	27.3	45.4	63.3	80.6
62H	0.8	42.9	61.7	68.5	83.0
62J	0.5	28.2	57.1	71.3	82.7
63B	0.5	35.0	55.1	87.7	111.7
63D	-0.6	38.6	61.2	75.1	97.0
63E	-8.3	31.3	64.4	92.7	114.3
63G	-0.5	29.4	53.7	81.5	98.0
63H	13.6	43.7	65.2	88.9	110.0
63J	2.7	37.0	61.7	83.0	96.2
63N	-21.5	35.0	57.1	78.0	115.2
63S	-3.1	28.2	63.1	84.2	86.6
63T	-6.7	31.3	62.3	90.4	114.3
63W *	-0.6	32.3	60.5	85.5	102.7
63Y	6.2	35.5	62.0	88.4	105.6
64C	22.7	48.8	52.2	81.9	79.6
65B	2.9	27.3	49.0	61.5	72.7
65D	16.0	29.3	66.2	70.8	85.5
65E	-13.6	23.3	41.3	69.2	87.7
65F	-22.3	24.5	43.7	70.9	77.3
65G	-9.9	32.8	43.3	73.5	85.2
65H	-3.4	29.5	58.2	77.1	54.2
65J	12.7	29.4	51.7	68.3	81.5
65K	-8.8	25.4	47.2	67.0	93.3
67G	-12.2	19.0	53.7	70.5	103.0
67H	-24.4	21.6	49.0	80.6	110.0
67N	-24.0	8.5	46.9	80.5	107.7
67R	-25.7	20.8	47.3	82.7	115.2
67S	-22.3	20.8	52.2	91.7	108.7
67T	-14.8	18.7	48.3	83.7	114.3
67U *	-17.9	23.0	56.6	93.6	119.7
67V	-14.7	21.1	51.7	70.5	105.6
67X	-16.3	25.4	50.7	80.6	120.6
67Y	-24.0	21.7	48.9	85.5	113.8

* One of 12 MOS assigned actual ratio scales

MOS	Percentile 10	30	50	70	90
68B	-25.7	9.8	41.3	80.3	115.2
68D	-25.4	22.7	42.0	77.3	108.7
68F	-24.7	23.2	45.6	77.0	108.2
68G	-16.4	20.5	46.4	81.5	102.1
68H	-14.7	19.0	49.6	79.2	105.6
68J	-16.3	21.6	57.9	82.6	115.2
68M	-13.3	21.7	46.9	83.0	113.8
71C	-0.6	22.7	47.1	56.9	97.0
71D	-10.0	29.5	48.3	70.6	88.2
71G	-2.8	33.4	53.7	74.8	90.8
71L	0.6	27.3	54.3	70.8	86.8
71M	8.5	33.1	61.7	66.2	83.0
71N	-10.8	28.2	50.5	86.6	115.0
71Q *	-9.9	16.9	36.7	51.4	62.6
71R	-4.0	27.3	43.7	59.7	67.0
72E	-6.3	30.5	63.0	80.3	101.9
72G *	-5.6	31.1	55.1	82.6	98.8
72H	-19.3	24.5	35.2	73.0	105.6
73C	-6.3	37.1	49.3	78.0	101.9
73D	-22.3	20.8	38.6	68.9	89.1
74D	-5.6	18.9	47.1	79.6	105.6
74F	-16.4	18.7	44.4	70.6	104.5
75B	-0.1	24.1	56.2	72.8	85.9
75C	1.7	25.3	55.8	79.2	103.0
75D	0.6	23.5	59.7	67.0	86.8
75E	4.6	25.5	61.7	85.5	93.5
75F	-2.9	30.5	47.3	80.3	105.0
76C	-3.4	33.1	56.2	81.5	77.1
76J	-7.5	23.2	55.8	66.3	93.1
76P	-4.0	27.3	54.3	76.6	88.9
76V	-3.2	40.9	55.2	78.0	90.8
76W	15.5	39.2	59.1	73.5	82.7
76X	6.4	28.2	52.2	79.6	91.7
76Y	-5.0	33.1	66.4	81.5	92.7
81B	2.7	27.4	51.0	68.5	93.5
81C	0.5	23.3	55.1	78.0	98.9
81E	-6.7	22.3	60.2	66.4	62.3
81Q	1.7	35.5	59.9	81.5	110.8

* One of 12 MOS assigned actual ratio scales

MOS	Percentile 10	30	50	70	90
82B	-1.7	31.0	49.0	72.7	86.8
82C	-5.1	37.5	51.7	86.1	108.2
82D	-11.2	19.8	53.1	78.0	96.2
83E	-0.6	26.4	55.7	68.9	77.3
83F *	-6.0	23.4	37.8	54.7	67.7
84B	-0.5	31.4	49.6	70.5	86.1
84C	-4.0	21.6	41.9	52.5	59.7
84F	0.8	25.5	57.3	63.9	70.8
91A	-4.0	17.6	52.5	76.6	100.2
91D	-19.6	12.3	42.9	68.5	104.8
91E	-6.3	25.8	49.3	78.0	93.1
91F	-5.6	22.7	42.0	75.1	91.7
91G	-13.2	20.5	46.4	58.2	79.3
91H	-2.8	23.2	45.6	72.6	88.4
91J	7.3	13.6	41.9	65.2	72.7
91L	0.8	23.6	53.1	73.1	88.1
91N	-21.5	23.3	32.8	75.7	101.9
91P	-13.6	24.5	50.5	81.9	91.7
91Q	-18.0	15.2	44.4	77.1	88.2
91R	-11.3	17.6	56.1	63.3	84.7
91S	-0.5	29.4	53.7	68.3	95.6
91T	-1.7	23.5	49.0	54.3	65.2
91U	-17.5	14.1	48.9	73.1	101.9
91V	-13.6	23.3	45.3	75.7	101.9
91W	-13.2	24.1	46.4	72.8	102.1
91Y	-5.6	10.7	42.0	70.9	89.1
92B	-2.8	29.4	55.8	74.8	103.0
92C	4.0	29.4	43.5	64.1	83.8
93D	-19.0	13.6	50.7	86.8	97.9
93E	-4.0	19.6	50.7	74.6	88.9
93F	0.8	27.4	61.7	73.1	101.9
93H	-14.8	17.0	48.3	70.6	104.5
93J	-17.1	14.8	49.6	72.6	105.6
93P	-11.3	21.6	54.3	80.6	107.5
94B	2.7	27.4	63.9	85.5	90.8
94F	3.7	32.8	51.3	73.5	82.7
95B	-8.2	38.6	63.1	84.2	108.7

* One of 12 MOS assigned actual ratio scales

MOS	10	30	50	70	90
			Percentile		
96B	-2.9	23.3	49.3	80.3	115.2
96D	-8.2	16.9	43.7	81.9	99.8
96F	-8.3	22.3	52.2	64.4	90.4
96H	-5.1	31.4	49.6	77.0	110.8
96R	-3.2	27.4	57.3	78.0	93.5
97B	-19.0	11.5	45.4	82.6	107.5
97E	-19.6	14.1	38.9	78.0	113.8
97G	-9.9	28.2	51.3	73.5	87.7
98C	-13.6	20.8	48.8	75.1	125.3
98G	-11.6	22.3	42.5	81.5	109.3
98J *	-7.8	24.6	59.5	90.0	114.3

* One of 12 MOS assigned actual ratio scales